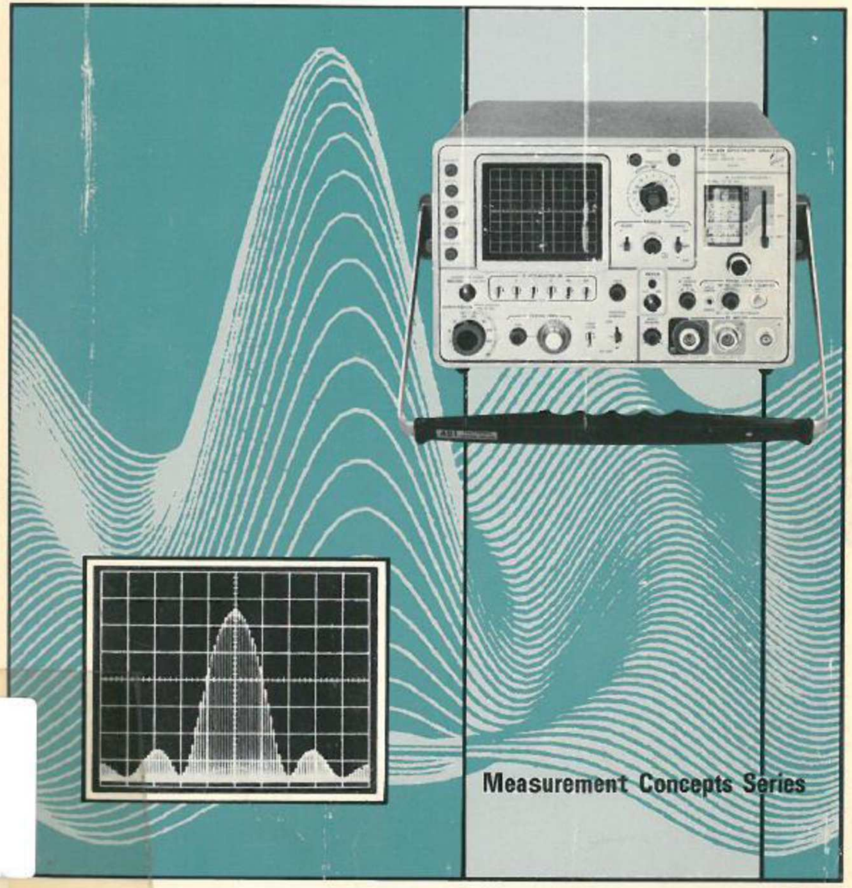
Spectrum Analyzer Measurements
theory and practice



SPECTRUM ANALYZER MEASUREMENTS

theory and practice

BY
MORRIS ENGELSON

MEASUREMENT CONCEPTS

SPECTRUM ANALYZER MEASUREMENTS

theory and practice

BY

MORRIS ENGELSON

MEASUREMENT CONCEPTS

FIRST EDITION, FIRST PRINTING, FEBRUARY 1971
062-1334-00

CONTENTS

PREFACE *1*

PART I **MEASUREMENT THEORY** *3*

1 SPECTRUM ANALYZERS *5*

2 SPECTRUM THEORY *17*

3 FOURIER ANALYSIS *43*

4 MODULATION THEORY *85*

5 THE SWEEPING-SIGNAL SPECTRUM ANALYZER *111*

PART II **MEASUREMENT PRACTICE** *131*

6 THE MEASUREMENT PROBLEM *133*

7 AMPLITUDE MODULATION *157*

8 FREQUENCY MODULATION *167*

9 PULSED RF *191*

10 MISCELLANEOUS APPLICATIONS *219*

11 DEFINITIONS OF TERMS *245*

APPENDIX *263*

BIBLIOGRAPHY *285*

INDEX *289*

PREFACE

This book is primarily concerned with the problem of measurements in the frequency domain by means of Spectrum Analyzers. Thus circuit design or construction details are not considered. Basic system parameters are, however, discussed in some detail since these have a direct bearing on the interpretation of measurement data. Two types of signals are treated in detail: Those composed of discrete or line spectra and those composed of continuous or dense spectra. Continuous wave (CW) or sinusoidal amplitude modulation (AM) is an example of the former, while pulsed-RF is treated as the latter. The third class of signals comprising random variables and requiring statistical methods are excluded from the detailed discussion, though some applications are included.

The discussion follows a dual approach: Part I is a mathematical-process-oriented approach while Part II applies the theory of Part I to specific measurement problems.

Those desiring to avoid mathmatical complexities can go directly to Part II where the basic relationships are stated without proof or explanation. Those who wish a somewhat deeper understanding should read selectively in Part I where the more abstruse material has been relegated to the appendices. Finally, those who need a thorough grounding in the subject, and this is unavoidable for those who wish to use Spectrum Analyzers in new and as yet untried areas, will hopefully find the proofs and references of interest.

PART I

MEASUREMENT THEORY

1

SPECTRUM ANALYZERS

frequency versus time

All electronic signals, indeed all natural phenomena, can be described either as a function of time or of frequency. When a phenomenon is cyclical, having a definite periodicity, the basic relationship between frequency and time interval is fairly simple, one being essentially the inverse of the other. In the case of random phenomena, one has to use statistical methods, but the concept of the duality of time and frequency is still useful. The concept of *frequency* as considered here presupposes *time* duration[1] — where time is a basic property of the universe we live in and frequency is related to time through the cyclical or periodic nature of the phenomena under discussion.

spectrum analyzer defined

Just as the oscilloscope is an instrument whose basic function is to display the time characteristics of phenomena, so is the spectrum analyzer an instrument whose function is to display the frequency characteristics of phenomena. The basic definition of a spectrum analyzer as found in Chapter 11 is: "A device which displays a graph of relative power distribution as a function of frequency, typically on a cathode-ray tube or chart recorder."

---

[1] One could consider frequency in more general terms – i.e., one could say that a topographical distribution of hills is cyclical with x hills per mile. Time need not enter such a discussion.

time or frequency domain

It should be recognized that the two descriptions of the same phenomena — *time domain* for the oscilloscope and *frequency domain* for the spectrum analyzer — are not independent. If one of the two is known, the appropriate mathematical rules or equations lead to the other. The question of which description, time or frequency domain, is the more basic is difficult to answer. One can argue that time is the basic natural phenomenon and that the frequency concept is derived from it since a universe without time duration is inconceivable to us. It can, however, also be argued that it is the periodicity of natural phenomena that makes for the thing we call time. Certainly in modern physics, such as relativity theory, it is considered that time stems from the existence of matter and hence movement and periodicity and that time without matter (Newton's *absolute time*) has no meaning.

time is the independent variable

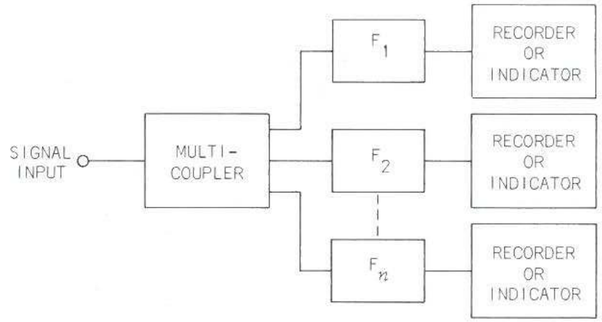
For the purposes of this discussion it is convenient to consider time as the basic concept and frequency derived from it. This is because oscilloscopes are constructed to display an enhanced (amplified, and/or sampled) version of the incoming signal, whereas spectrum analyzers are constructed to obtain the frequency-domain characteristics of the incoming signal by computation or other analog operation performed on a time-domain input. Thus, while an oscilloscope is generally recognizable as such from its block diagram, a spectrum analyzer may be difficult to recognize, since the computation function or analog operation can be performed by diverse means.

Some of the methods that can be used to obtain a frequency-domain presentation are described in the following.

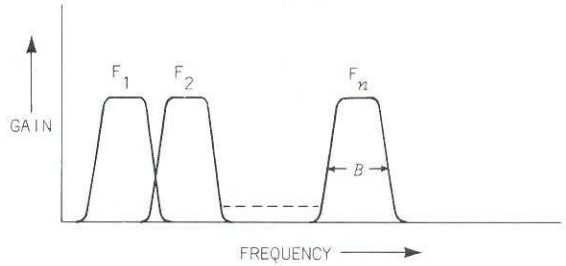
CONSTRUCTING A SPECTRUM ANALYZER

computer analysis

One method of performing spectrum analysis is to program a computer to perform the appropriate computation for going from the time to the frequency domain. With the recent advent of the time-saving Fast Fourier Transform technique, this method of spectrum analysis can be quite attractive. Nevertheless, the computer programmed as a spectrum analyzer is still a rarity which is used for specialized applications only.



(A) CONTIGUOUS FILTER BANK, BLOCK DIAGRAM.



(B) CONTIGUOUS FILTER BANK, FREQUENCY CHARACTERISTICS.

Fig. 1-1.

frequency filters

Besides the computer method, the frequency composition of a complex signal can also be obtained by separating the several components in a contiguous bank of filters. Such a system of contiguous-filter passbands can be constructed from real physically existing filters. Figs. 1-1A and 1-1B are the block diagram and frequency characteristics of such a system. The composite time-domain signal is fed to the multicoupler which distributes it equally between the several filters. Each filter will respond only to inputs within its passband. Hence, by observing the amplitudes of the outputs of the various filters one can determine which frequencies and what amplitude levels are present in the composite input signal.

resolution

dispersion

The narrower the filter bandwidth B, the better our ability to determine the precise frequencies of the signal components. This discrimination between signals having closely spaced frequencies is called *resolution*. The narrower B, the better is our resolution. This is analogous to the resolution of a microscope, where an improvement in resolution refers to an increased ability to separate visually several small particles. The range of frequencies that can be analyzed is of course the total frequency band covered by the set of contiguous filters. If there are a total of n filters each having a bandwidth B, then the total frequency range is simply the product nB. This is sometimes referred to as the *dispersion*, a word borrowed from optics. As the filter bandwidth is decreased in order to improve resolution, it becomes necessary to increase the number of filters (n) correspondingly if the dispersion is not to decrease. Thus, the number of filters and indicators in such a system can get very large. One way to reduce the cost of such a system is to use one recorder, or indicator, which is commutated between the various filters, as depicted in Fig. 1-2. Here, instead of determining frequency by which indicator shows an output, we determine frequency by correlating the time of the output with the sequence of the commutator. For example, if the commutator is connected to each filter for half a second and it takes a half-second travel time to go from filter to filter; an output at 10 to 10.5 seconds after start means a signal corresponding to the frequency of the tenth filter. With appropriate recorder speed, for example one inch per second, we can transform the recorder time axis into a frequency axis. In the previous example a signal indication positioned 10 inches from the start means a signal at the frequency of the tenth filter.

Superficially, the systems shown in Fig. 1-1A and Fig. 1-2 seem to accomplish the same thing. There is, however, one major difference between them. Whereas the first system will, at least in theory, show all signals no matter how short their duration, the second system is limited in this respect by the speed of the commutator and the time constants or memory associated with the system. Thus, the convenience of a reduction in system complexity is paid for by the loss of some capability.

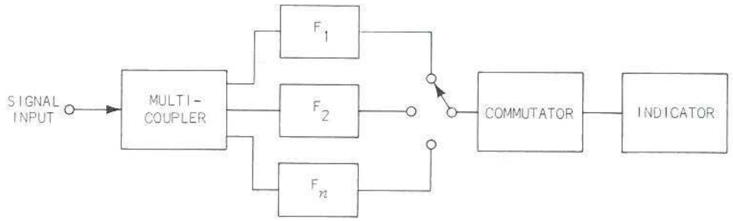


Fig. 1-2. Filter bank system using commutator.

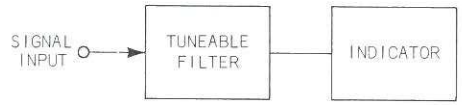


Fig. 1-3. Tuned-filter spectrum analyzer.

tuned filter

The system shown in Fig. 1-2 is still quite cumbersome since it may require hundreds of filters to obtain the desired resolution and dispersion. However, since we no longer utilize our filters on a continuous basis, why not use a single filter whose center frequency is switched or tuned in lieu of the commutator. This results in the system depicted in Fig. 1-3. The characteristics of such a system are best understood by considering the relationships shown in Fig. 1-4. Here a filter having a bandwidth B is assumed to tune over the frequency range f_1 to f_3 during the time interval T.

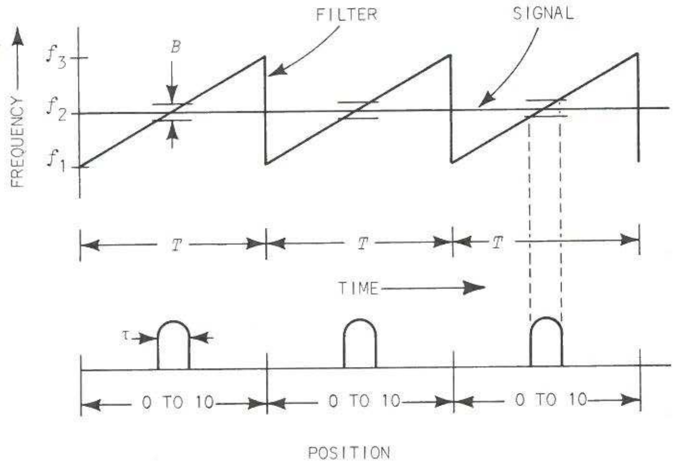


Fig. 1-4. Frequency/time-position relationships for tuned-filter system.

During the same time interval, the indicator (paper chart recorder or CRT) changes position from zero to ten divisions. For example, zero position corresponds to f_1, 10 divisions corresponds to f_3 and 5 divisions corresponds to the center frequency between f_1 and f_3, namely $\frac{f_1 + f_3}{2}$. In addition, Fig. 1-4 also shows a signal at frequency f_2. The signal is shown as a straight line in the time-frequency diagram, meaning that it has constant frequency as a function of time. The effect of the tuning or sweeping filter intercepting the signal is indicated by the pulses on the horizontal position scale. The width of the pulses traced by the indicator is $\tau = \frac{B}{f_3 - f_1} T$, namely, the time that the signal frequency is within the passband of the filter.

This system is relatively simple and compact, but there are practical difficulties. Problems stem from the present state of the art in electronically tuneable-filter construction. These generally have much wider bandwidths than is desired for most applications. Thus, spectrum analyzers of this type have somewhat limited utility. This brings us to our final configuration, the sweeping superheterodyne system.

sweeping signal

A major point to be recognized in the system of Fig. 1-4 is that the transformation from time domain to frequency domain is accomplished by the relative translation or movement in frequency between the filter and the signal. The emphasis should be on the word *relative*, meaning that it does not matter whether it is the filter or the signal frequency that is changing or translating. Thus, one should be able to obtain the same end result as that of Fig. 1-4 by using a stationary filter and a translating, or to use the more common name *sweeping*, signal. The time/frequency-position relationships for such a system are shown in Fig. 1-5. Here the filter is shown as having a passband (B) and a constant unchanging center frequency (f_2). When we have a signal whose frequency falls within the passband of the filter, there is an output. The result is the pulse on the position scale whose position corresponds to a frequency, f_2, and where the pulse width $\tau = \frac{B}{f_3 - f_1} T$ is a measure of system resolution.

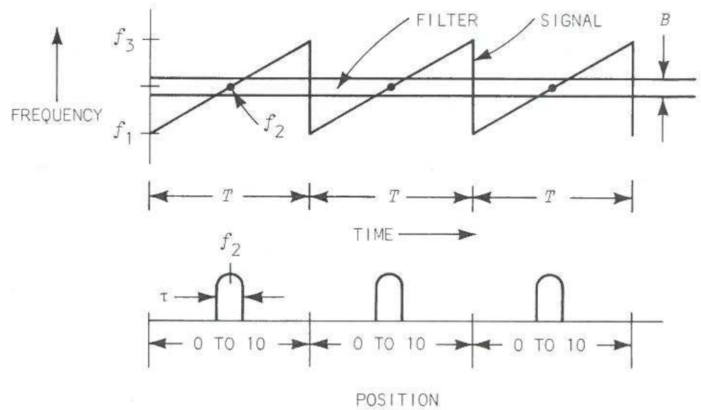
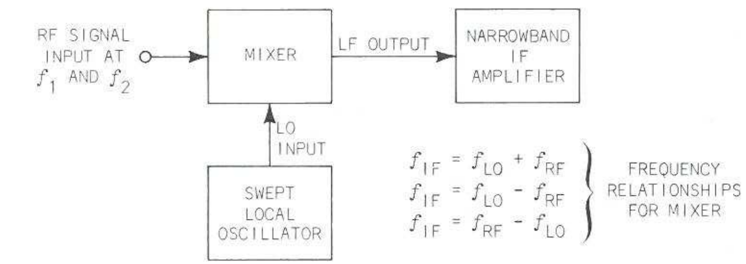
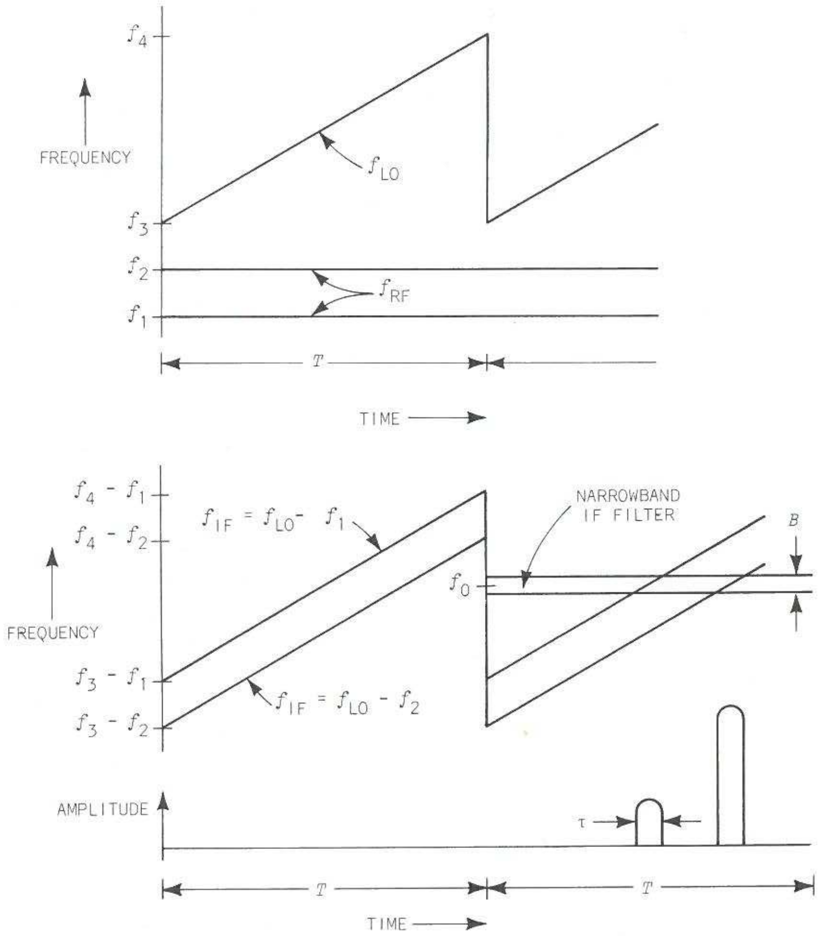


Fig. 1-5. Frequency/time-position relationships for swept-signal system.



(A) BASIC SYSTEM



(B) FREQUENCY-TIME RELATIONSHIPS USING $f_{IF} = f_{LO} - f_{RF}$ MIXING SYSTEM.

Fig. 1-6. Superheterodyne system.

THE SWEEPING-SIGNAL SYSTEM[2]

mixer

The superheterodyne or sweeping-signal system is based on the use of a *mixer*. For the present we shall consider the mixer as an idealized three-terminal black box. The three terminals provide for a signal input, local oscillator input and IF output. Our idealized mixer produces an IF output which has the amplitude characteristics of the signal input and whose frequency characteristics consist of an algebraic combination (sum or difference) of the frequency characteristics of both inputs. We produce what amounts to signal sweeping by sweeping the local-oscillator frequency which produces a swept IF output. Fig. 1-6 is a block diagram and frequency-time display showing the effect of multiple signal frequencies. Here we observe that for each signal frequency we generate a separate IF frequency sawtooth through one of the relationships:

$$f_{IF} = \begin{cases} f_{LO} + f_{RF} \\ f_{LO} - f_{RF} \\ f_{RF} - f_{LO} \end{cases}$$

Fig. 1-6B is drawn for the $f_{IF} = f_{LO} - f_{RF}$ relationship. As the IF frequency sawtooth passes by the filter, consisting of the narrowband IF amplifier of bandwidth B and center frequency f_0, we generate a pulse of width τ, as previously discussed. One such pulse is generated for every signal frequency present, with pulse height proportional to signal amplitude. In Fig. 1-6B, it was assumed that the amplitude of the signal at f_2 is larger than that at f_1.

[2] Detailed characteristics are described in Chapter 5.

TYPES OF SIGNALS

continuous wave

pulsed RF

line spectra

continuous spectrum

The examples considered thus far are based on the assumption that the input can be considered as consisting of several sinewaves. If we know the response of our system to a continuous-wave (CW) signal, we know all that's needed since the rest follows from simple superposition. This assumption is not always warranted. A case in point is pulsed RF. Here we obviously cannot obtain the output pulse shown in Fig. 1-6B if the input-pulse duration is less than the width τ and the response of the system to pulsed inputs is much more complicated than that for CW type inputs. In the CW case, the result is a pulse of width τ which is dependent on the resolution bandwidth, the dispersion and the sweep time, while in the pulsed-RF case we have a pulse considerably narrower than τ, where the pulse width is determined by the signal characteristics rather than the spectrum-analyzer parameters. The former type of signal is considered as consisting of *line spectra* — discrete CW components; the latter type of signal is described by a *continuous spectrum* or a *dense spectrum.* For the moment it is helpful to recognize one basic difference: In the CW case, the final result is a pulse tracing out the shape of the resolution filter which can be considered as the steady-state response of the narrowband IF amplifier; in the continuous-spectrum case, the narrowband amplifier has to respond to a fairly narrow pulse resulting in a transient rather than steady-state response. The significance of this and other differences between these two classes of signals is considered in Chapter 5.

Now that the reader has an idea of what we mean by the words "Spectrum Analyzer" and how the device operates, we shall put aside the hardware and proceed with a theoretical discussion on the frequency-domain characteristics of signals. The practical aspects of measurements will be considered in Part II of this volume.

numerical examples

Referring to Fig. 1-6A, assume that the following numbers apply:

1) The narrowband IF

$$f_0 = 75 \text{ MHz}$$
$$B = 100 \text{ kHz}$$

2) The swept local oscillator

$$f_3 = 270 \text{ MHz}$$
$$f_4 = 280 \text{ MHz}$$
$$\text{Sweeptime } T = 10 \text{ ms}$$

Consider the response to a CW signal at 200 MHz.

The signal can combine with the local oscillator as follows:

$$f_{IF} = \begin{cases} f_{LO} + f_{RF} \\ f_{LO} - f_{RF} \\ f_{RF} - f_{LO} \end{cases}$$

In our case, we are only interested in the results where $f_{IF} = f_0 = 75$ MHz. This means that we are interested in the relationship $f_{LO}\text{-}f_{RF} = f_{IF}$, since 275 - 200 = 75. This happens when the local oscillator is 275 MHz, which is the center of its sweep, 5 ms from sweep start. Actually we get an output not only at 75 MHz but at 75 MHz ±50 kHz since the bandwidth is 100 kHz. Hence the result is a pulse of width $\tau = \frac{B}{D}T = \frac{0.1}{280 - 270}10 = 0.1$ ms centered at 5 ms from the start of the sweep. Assuming that this pulse were displayed on a CRT having 10 horizontal divisions at 1 ms/div (same as the sweeping LO), we would have a pulse occupying 0.1 divisions. It is interesting to note that the pulse continues to occupy 0.1 divisions regardless of what the sweeptime T is, so long as the sweeping oscillator and the time base of the CRT are identical. Thus, if T is made 100 ms, τ becomes 1 ms and at 10 ms/div still occupies 0.1 divisions.

Consider now that the signal consists of two components — one at 200 MHz and the other at 202 MHz. The result is two output pulses, one at 5 ms from the start and the second at 7 ms from the start. These would appear on the CRT at 5 and 7 divisions respectively. The relative amplitudes of these pulses would be in the same proportion as the relative amplitudes of the signal components. Proceeding in similar fashion, we see that the horizontal CRT scale can be considered as a frequency scale where the left-hand edge represents 195 MHz and the right-hand edge represents 205 MHz.

Consider now that the signal has a third component at 350 MHz. This too will appear on the CRT via the relationship $f_{RF} - f_{LO} = f_{IF} = 350 - 275 = 75$. Yet the CRT has not been calibrated for it, since we've assumed that our frequency base is 195-205 MHz. Such a signal, which does not conform to the frequency calibration of the CRT, is called a *spurious response.* There are many types of spurious responses. This particular spurious response is called the *image.* Let us now go back to our original signal at 200 MHz and see what happens to it as a function of the spectrum-analyzer control settings. As previously determined, this signal appears on the CRT as a response 0.1-div wide located in the center of the CRT with an amplitude which is proportional to the input level.

spurious response

image

1) Changing the sweeptime: As previously determined, this should have no effect on the appearance of the pulse. Thus, the width of the signal pulse is a true measure of relative resolution since it is, at least in theory, dependent only on resolution and dispersion. In actuality, if the sweeptime is reduced too much, there will be major changes in what appears on the CRT. This aspect will be discussed in the section on spectrum-analyzer limitations.

2) Changing the local-oscillator center frequency: Let the sweeping local oscillator operate from 271-281 MHz. Thus, 275 MHz occurs four-tenths from the beginning and our signal pulse will move from the fifth to the fourth graticule line on the CRT. This is because the CRT frequency base has now changed from 195-205 MHz to 196-206 MHz.

3) Changing the sweeping-oscillator sweep width or dispersion: Let the sweeping oscillator operate from 272.5-277.5 MHz for a total excursion of 5 MHz. The pulse width is $\tau = \frac{B}{D} T = \frac{0.1}{5} T$ seconds which occupies a physical distance of $\frac{0.1}{5} T$ seconds $\times$ 10 CRT divisions = 0.2 div, or twice the previous width. The frequency base of the CRT has likewise been changed from 195-205 MHz to 197.5-202.5 MHz.

4) Changing the resolution bandwidth: Let the resolution bandwidth B = 50 kHz. The only effect is to reduce the signal pulse width to $\tau = \frac{0.05}{10} T$ seconds or 0.05 divisions wide on screen.

2

SPECTRUM THEORY

TIME AND FREQUENCY DOMAIN

Any motion, or to use the electronic term *waveform*, which repeats itself as a function of time, is called *cyclical.*

When the waveform repeats in equal intervals of time, it is considered *periodic.*

definitions

When a waveform is generated by the retracing of the same path, such as the back-and-forth motion of a pendulum or the back-and-forth transfer of charge in an *LC* circuit, it is called *oscillatory.*

One complete *oscillation* means that a round trip is completed; e.g., from A to B and back to A again.

The *period (T)* of the oscillation is the time required to complete one oscillation.

The *frequency (f)* is the number of oscillations per unit time, i.e., $f = \frac{1}{T}$, $T = \frac{1}{f}$.

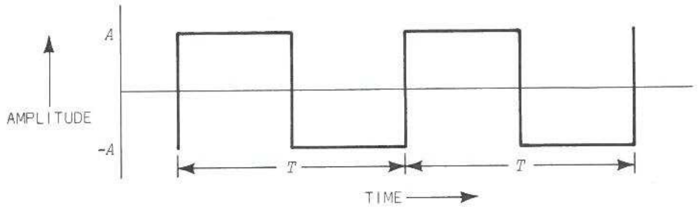


Fig. 2-1. Periodic waveform, time domain.

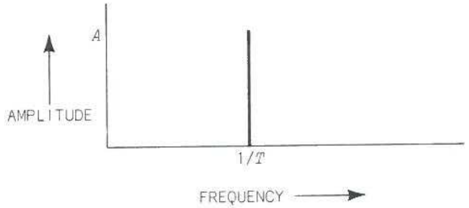


Fig. 2-2. Periodic waveform, frequency domain.

squarewave

With the basic definitions behind us, let us now consider a periodic waveform such as the squarewave shown in Fig. 2-1. This waveform can be described as having a period T or a frequency $f = 1/T$; either of these statements combined with the statement that the amplitude is A gives us a complete description of the squarewave. However, when comparing our description with the graphical representation, it is observed that the description in terms of the period T is easier to use since the description in terms of frequency is not applicable to the coordinate system of the graph without mathematical computation. Yet it is often more meaningful to describe a phenomenon in terms of frequency rather than time duration. This leads to the desirability of producing a graphical representation of the squarewave as a function of frequency rather than time. Such a graphical representation is Fig. 2-2. Fig. 2-2, however, only conveys the appropriate information to those who know the conventions used. Thus, we must specify that our basic waveform is a squarewave, not a triangle or trapezoid or sinusoid. We must also be aware that by showing the frequency domain representation as positive it is meant that the squarewave starts out with a positive rather than negative excursion.

Knowing the conventions, it is possible to represent all kinds of waveforms in the frequency domain by breaking these up into squarewaves. Such a representation is shown in Fig. 2-3. Here the complex pulse shape of Fig. 2-3C is represented as the sum of the two squarewaves shown in Fig. 2-3A and 2-3B. These squarewaves are in turn represented in the frequency domain by Fig. 2-3D. As far as information content is concerned, Figs. 2-3C and 2-3D are completely equivalent, each being derivable from the other.

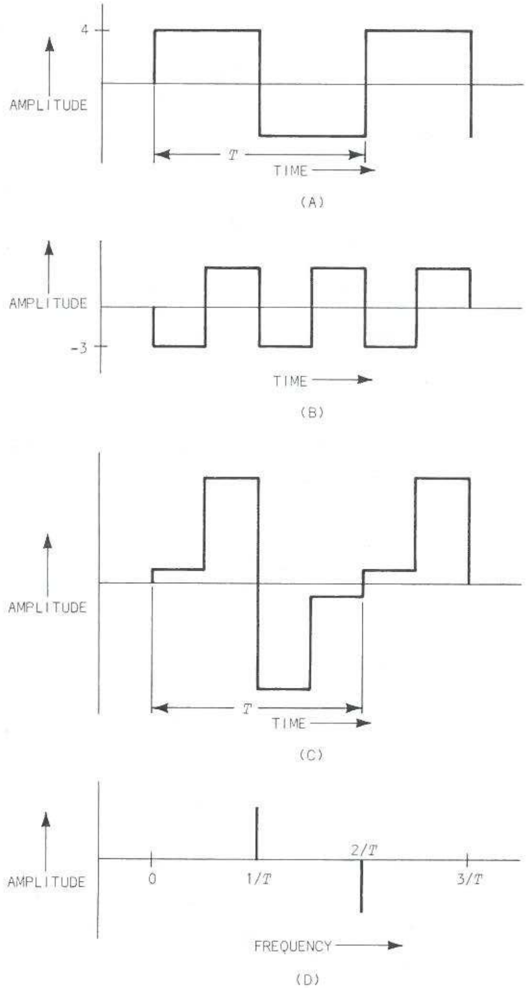


Fig. 2-3. Time- and frequency-domain representation using squarewave as basic frequency function.

Now, it should be recognized that there is nothing sacred about using squarewaves as the basic waveform. For all we know, triangles or trapezoids or sinusoids might be much more convenient. The next order of business is to consider which properties are desirable for the basic waveform.

ORTHOGONAL FUNCTIONS

functions

orthogonality

The most important requirement of the basic waveform is that as many as possible (preferably all) other waveforms should be disassociable into a combination of basic waveforms. Though many types of waveforms, or to use the proper mathematical terminology *functions*, can be used as the basic function, it can be shown[1] that sets of functions which possess the property of *orthogonality* fulfill the above requirement best. Let us now consider the meaning of the word *orthogonal*.

The word orthogonal comes from the words *orthos* meaning right and *gonia* meaning angle. In ordinary usage the word is defined as pertaining to right angles. The mathematical meaning is more precise but based on the same classical roots. It is based on the fact that when two lines or planes are at right angles to each other, the projection of one onto the other is zero, as illustrated in Fig. 2-4. Specifically, a set of functions is orthogonal when the integral between specified limits of the product of any two functions is zero.

In mathematical notation when:

$$\int_a^b f_m(x)\, f_n(x)\, dx = 0, \text{ when } m \neq n,$$

then $f_1(x), f_2(x) \ldots f_m(x), f_n(x)$ form an *orthogonal* set of functions.

[1] See, for example, Whittaker & Watson's *Modern Analysis* for a detailed mathematical exposition on the expansion of functions in infinite series using both orthogonal and nonorthogonal functions.

One can interpret orthogonality as a geometrical condition by considering that the result of an integration is the area under the curve bounded by the function being integrated. Thus, when a set of functions is orthogonal, what we mean is that the area under the curve generated by the product of any two functions, except the function times itself, is zero.

A physical example of orthogonal functions is a set of three mutually perpendicular vectors. For this case, according to the equation, the projection of any one vector upon any other vector is always zero.

There are many sets of orthogonal functions. For example: Sinewaves, Bessel functions, and the series composed of $1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x, \ldots$ taken between the limits of ± 1, are all orthogonal.

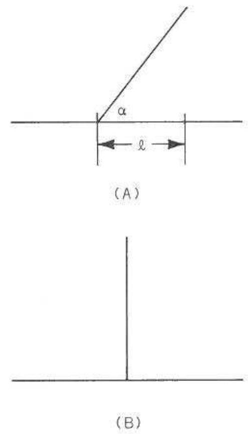


Fig. 2-4. Projection is ℓ when $\alpha \neq 90°$; projection $f(x)$ is zero when $\alpha = 90°$.

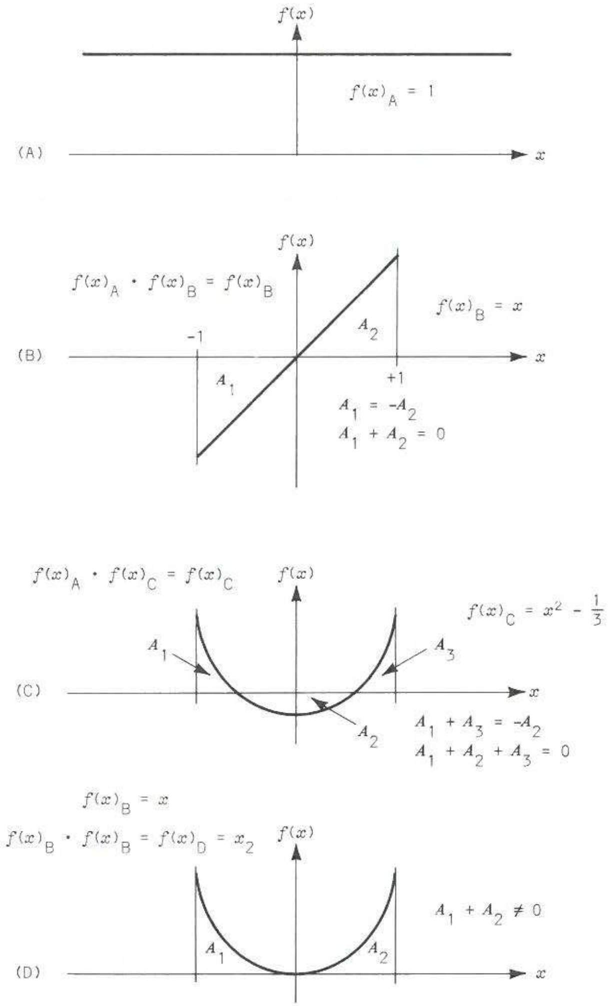


Fig. 2-5. Geometrical interpretation of the orthogonal series: $1, x, x^2 - \frac{1}{3}$, ... the integral between the limits of the product of any two terms is zero as demonstrated in (A), (B) and (C), while the integral of the product of any term times itself is not zero, as illustrated in (D).

A geometrical interpretation of this series is shown in Fig. 2-5.

As previously indicated, a series of orthogonal functions serves our purposes best. This is because almost any function[2] defined over a specific interval, such as $+\pi$ to $-\pi$, can be expanded in almost any set of orthogonal functions. Thus, from the point of view of what can be done theoretically, Bessel functions, which we shall discuss in connection with frequency modulation, are just as good as sines and cosines. The question of which orthogonal set of functions to use must, therefore, be settled on the merits of practicality rather than theory. This is considered in the next section.

THE PROPERTIES OF SINEWAVES

Among the various sets of orthogonal functions, that consisting of sines and cosines comes closest to describing the behavior of physical systems and is one of the easiest to manipulate mathematically as well.

The sine functions have the following important properties:

sine functions

1) The sine function is generated in connection with motion around a circle, as will be demonstrated later. Since much of our machinery is based on circular motion, the sine function can be used to describe physically existing situations.

2) Most physical processes which are undulatory are also periodic. Unlike the orthogonal Bessel functions, for example, which are undulatory but not with a constant period, the sine functions are periodic.

---

[2] There are some unimportant exceptions.

3) Many diverse phenomena such as the oscillation of a weight on a spring, the swing of a pendulum, or the oscillation of current in an *LC* circuit are basically sinusoidal. In particular, the fact that the natural behavior of electrical circuits is sinusoidal is of the greatest importance in our choice of sinewaves as our basic waveform[3].

4) Sinewaves possess the remarkable mathematical property: The basic description remains invariant under various mathematical transformations. For example, except for a change in phase, a sinewave remains a sinewave with integration or differentiation.

Based on reasoning such as above, one comes to the conclusion that, among the various sets of orthogonal functions, sinewaves are both easier to manipulate and come closer to describing physical processes than any other type of function. Hence, when describing complex waveforms by breaking these up into a sum of more elementary waveforms, the sinewave is chosen as the basic waveform.

Let us now consider some of the terminology and properties associated with sinewaves.

sinewave properties

The behavior of many physical systems is described by a differential equation of the form

$$\frac{d^2y}{dt^2} + \omega^2 y = 0,$$

a solution of which is

$$y = A \cos \omega t + B \sin \omega t.$$

[3] Actually, the damped rather than the constant-amplitude sinusoidal oscillation is the natural response of real networks, since all physical networks contain some losses. However, the undamped sinewave is so much easier to manipulate that it has become the universal choice as the basic waveform.

In electronics such an equation is basic to *LC* circuits

$$\frac{d^2Q}{dt^2} + \left(\frac{1}{\sqrt{LC}}\right)^2 Q = 0,$$

where L, C and Q are inductance, capacitance and charge respectively.

The solution is in the form of a sinusoidal oscillation at angular frequency $\omega = \frac{1}{\sqrt{LC}}$.

The functions $y = A \cos \omega t$ or $y = B \sin \omega t$ are often referred to as circular trigonometric or just *circular functions.* These are also sometimes connected with the words *simple harmonic motion,* since these functions describe the simple harmonic motion of a point around a circle, as well as other physical phenomena. We shall simply use the words *sinewave* or *sinusoidal* whenever possible. It should be understood that the general word *sinewave* refers to both $\sin \theta$ and $\cos \theta$.

The general expression for a sinewave is

$$y = A \sin \theta .$$

terminology

A represents the *amplitude* while θ represents the *angle.* In electrical problems, θ is usually replaced by the time-dependent expression

$$\omega t + \alpha ,$$

where ω is the radian frequency or *angular velocity*, the combined quantity $(\omega t + \alpha)$ is the *phase*, while the fixed angle α is the *initial phase.* t is *time duration*, usually counted in seconds.

The angular velocity is usually broken up into the expression

$$\omega = 2\pi f$$

where f is the *frequency*, namely how many cycles of the phenomenon occur in one second. The *period*, which is the inverse of frequency, is

$$T = \frac{1}{f}.$$

The generation of sinewaves is most easily visualized in connection with motion around a circle as discussed in the following.

sinewave generation

Fig. 2-6 represents a point (P) moving around a circle (radius A) in a counterclockwise direction with angular velocity ω radians per second. The figure also shows the curves generated by the projection of the moving point on the horizontal x axis and the vertical y axis. These curves trace the cosine and sine respectively. In the construction we have assumed that the initial position of P is that marked 0, if this were not the case, we would add the starting angle α as an initial phase angle.

Such a diagram is very useful since one can deduce many of the characteristics of sinewaves directly from it. For example, when the angle (θ) is zero the horizontal projection is A while the vertical is zero. Likewise, when the angle is $\pi/2$ radians (90°), the horizontal projection is zero while the vertical is A. Going further one observes that the two projected lengths of the radius A are equal when the point P is midway between zero and $\pi/2$. At this time the x and y projections are $A/\sqrt{2}$. In such a way one can construct a table of values for the sine and cosine such as that in Table 2-1. The table shows the angle θ in degrees, of course we could just as well have used radians since one full circle is 2π radians or 360°.

From Fig. 2-6 we observe that at 270° the point (P) is at position 3. This position can also be reached by moving 90° clockwise. So, if we designate counterclockwise rotation as positive and clockwise rotation as negative, we see that 270° is equivalent to -90° and vice versa. Using this type of notation, we can conclude with the help of Table 2-1 that the following relationships should hold:

$$\sin\theta = -\sin(-\theta)$$
$$\cos\theta = \cos(-\theta)$$

Many other relationships can be developed in a similar manner.

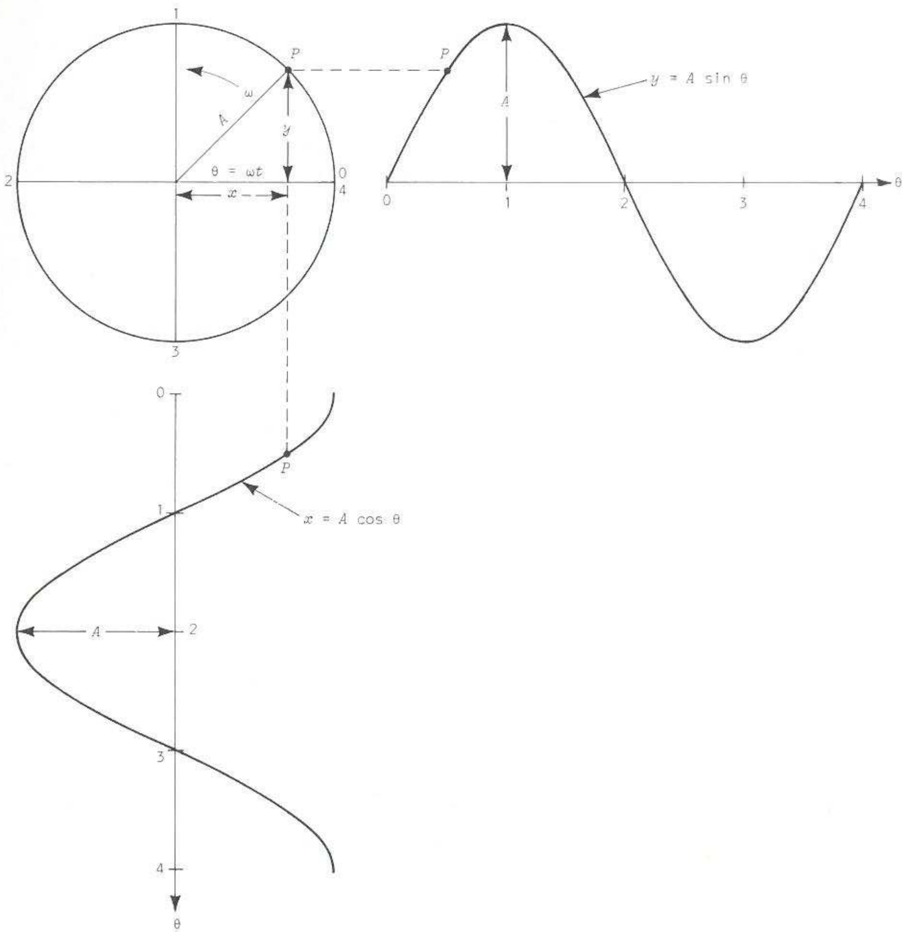


Fig. 2-6. Generating sinewaves by circular motion.

θ =	0	45°	90°	135°	180°	225°	270°	315°	360°
sin θ	0	$1/\sqrt{2}$	1	$1/\sqrt{2}$	0	$-1/\sqrt{2}$	-1	$-1/\sqrt{2}$	0
cos θ	1	$1/\sqrt{2}$	0	$-1/\sqrt{2}$	-1	$-1/\sqrt{2}$	0	$1/\sqrt{2}$	1

Table 2-1. Sine and cosine values.

The rotating point (P) starts at position zero at time $t = 0$ and starts moving counterclockwise. Eventually, (P) returns to its starting point (position 4); at this instant the elapsed time is some number, call it T. This time (T) is what we previously defined as the period. Likewise, the point (P) completes $f = 1/T$ complete trips around the circle every second. Since there are 2π radians in a circle, the angle that is covered in one second is $2\pi f$ (the angle per circle times circles per second). The angle swept to any arbitrary time t is merely the angle per second times the time in seconds or $\theta = 2\pi ft$. Thus our basic equations are

$$y = A \sin 2\pi ft,$$
$$x = \cos 2\pi ft$$

where $2\pi f = \omega$, the angular velocity, and f is the frequency as previously defined.

rotating vector

Besides the familiar notation, such as $x = \cos(\omega t + \alpha)$, used above, it is possible to represent sinewaves in diverse ways. One way, which was used above, is by means of the rotating vector[4] which is the radius of the circle in Fig. 2-6. The cosine function is represented by the projection of the rotating vector on the horizontal axis. The vector idea can be useful when dealing with the superposition[5] of several sinusoids. Thus, if we wish to find the sum of $A \cos \omega_0 t$ and $B \cos(\omega_0 t + \alpha)$, all we have to do is to construct a vector triangle with two vectors size A and B forming the angle α. The resultant vector (or phasor), rotating at angular velocity ω_0, is the solution to the problem.

two vectors

Strictly speaking, the single vector method is not quite correct since the rotating vector generates both a sine and a cosine function. In order to represent only one of these, it is necessary to use two vectors rotating in opposite directions as illustrated in Fig. 2-7. Here the two vectors have equal projections on the horizontal axis which represents the cosine, but equal and opposite projections on the vertical axis which represents the sine. Thus, if these vectors are half the amplitude of the single vector representation, we get, by addition, the same cosine function as before, but the amplitude of the sine is zero since the two vectors cancel each other.

---

[4] Such a vector is sometimes termed a *phasor.*

[5] Superposition is discussed more fully in the next chapter. For the moment it is sufficient to equate superposition with addition.

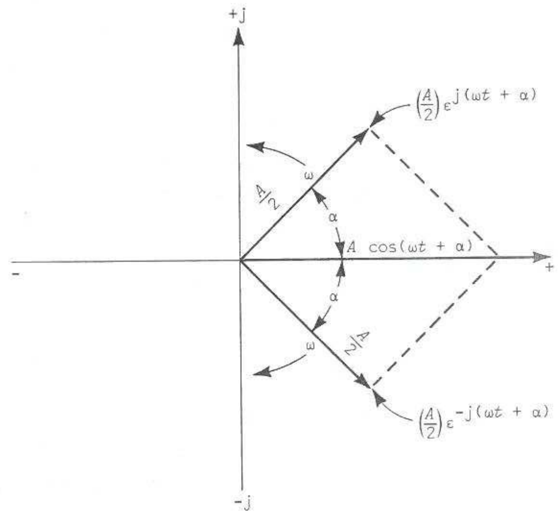


Fig. 2-7. Representing the cosine function by two counter-rotating vectors on complex plane.

complex notation

Such vectors can be represented in complex notation and manipulated in accordance with the rules of complex mathematics. Thus, in the complex plane, the horizontal axis is considered to represent real numbers while the vertical axis, labeled j, is considered to represent imaginary numbers. The word *imaginary* arises from $j = \sqrt{-1}$, which was once considered not to have any meaning and is just a name representing a mathematical expression; it has nothing to do with existence or nonexistence. Complex notation provides a simple mathematical expression for a complex combination of sines and cosines. Such notation is frequently used in Fourier series, hence our interest. In complex notation, a rotating vector is represented by the product of the amplitude and epsilon (the base of natural logarithms) raised to the power of the angle times j. In Fig. 2-7 the two counter rotating vectors are represented in complex notation by

$$\left(\frac{A}{2}\right) \epsilon^{+j(\omega t + \alpha)} \text{ and } \left(\frac{A}{2}\right) \epsilon^{-j(\omega t + \alpha)},$$

where the minus in front of the exponent on the second vector represents clockwise rotation, since counterclockwise is taken as the positive direction. The fact that such notation involves the use of negative frequencies should cause no concern. Remember that we are dealing with a mathematical notation and it is not necessary to ascribe physical reality to all the parts. Another way of looking at it is that the positive and negative aspects are just different notation standing for the same thing, since, as shown before, $\sin(\theta) = -\sin(-\theta)$ and $\cos(\theta) = \cos(-\theta)$[6].

trigonometric notation

The relationships between the complex notation and standard trigonometric notation can be derived from the rules governing the use of complex numbers. We shall skip a rigorous derivation, referring those interested to the references. The truth of the basic relationships can also be surmised with the help of a geometrical construction such as Fig. 2-7. Here it was shown that the sum of the two counter rotating vectors is the cosine function. Thus,

$$\frac{A}{2}\left[\epsilon^{j(\omega t + \alpha)} + \epsilon^{-j(\omega t + \alpha)}\right] = A \cos(\omega t + \alpha).$$

From geometrical reasoning such as this come the following results:

1) $\dfrac{\epsilon^{j\theta} + \epsilon^{-j\theta}}{2} = \cos\theta$

2) $\dfrac{\epsilon^{j\theta} - \epsilon^{-j\theta}}{j2} = \sin\theta$

3) $\epsilon^{j\theta} = \cos\theta + j\sin\theta$

These are not independent expressions, we can for example derive (1) from (3) thus:

$$\begin{aligned} \epsilon^{-j\theta} &= \cos(-\theta) + j\sin(-\theta) = \cos\theta - j\sin\theta \\ \epsilon^{j\theta} &= \cos\theta + j\sin\theta \\ \hline \epsilon^{-j\theta} + \epsilon^{j\theta} &= 2\cos\theta \end{aligned}$$

[6] For a more detailed discussion see, R. B. Marcus, "The Significance of Negative Frequencies in Spectrum Analysis" *IEEE Transactions*, EMC, Dec 1967.

sinusoid parameters

A sinusoid contains three basic parameters: amplitude, frequency and initial phase. Since a graphical representation in the frequency domain is essentially a two-dimensional process, only two parameters can be represented per graph. Thus it takes two graphs to define all three of the parameters. This can be done in two ways. One is to present a graph of frequency and initial phase, while the other way is to use the two-vector representation showing a negative as well as positive frequency. The trigonometric expression can be reconstructed from either graph. Fig. 2-8 shows both graphical representations. Though we shall use the complex notation, representing the two-vector technique, for analytical purposes, our graphical methods will be strictly of the single positive-frequency type. This is because in the majority of spectrum analyzer problems, phase is of no interest. We are thus able to represent all that is of interest in a single simple frequency-amplitude diagram. Here we have gone full circle back to Fig. 2-2, except that instead of squarewaves we shall utilize sinewaves as the basic function.

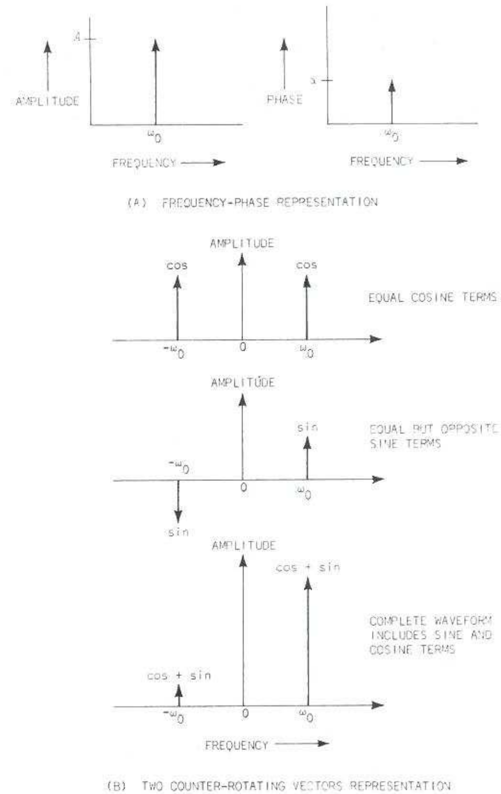


Fig. 2-8. Frequency-domain representations of sinusoid.

WHY FOURIER SERIES

Sines and cosines can be combined in many different ways to produce a final waveshape. Obviously, we have to settle on one particular method, and that is the Fourier series. Other mathematicians had used this series prior to Fourier. But it was Fourier who, in a series of papers starting in 1822, showed the universal applicability of the series that now bears his name.

The Fourier series is based upon the fact that any function, which meets the three conditions stated below, can be expanded in a series of sines and cosines. The so-called Dirichlet conditions, that a function should meet in order to be Fourier expandable, are:

Fourier expansion conditions

1) The function $f(t)$ must have only a finite number of maxima and minima for the interval of definition.
2) The function must have only a finite number of finite discontinuities.
3) If the function has infinite discontinuities, its integral must be convergent — i.e.,

$$\int_{-\infty}^{+\infty} |f(x)|\, dx < N \text{ (a finite number)}$$

These conditions are easily met in practice, since no physical circuit can produce an infinite discontinuity. Even when such discontinuities are approached, such as in an approximate impulse function, the third condition is always met. Thus the Fourier series is highly suitable to the solution of practical problems.

The basic Fourier series is of the form:

Fourier series

$$f(x) = \frac{1}{2}a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + (a_3 \cos 3x + b_3 \sin 3x) \ldots \quad (2\text{-}1)$$

It will be observed that equation (1) consists of a set of harmonically related terms where the lowest frequency terms ($\cos x$, $\sin x$) are considered the *fundamental* and all other terms are called the *harmonics.* A harmonic relationship means that all frequencies are integral multiples of the fundamental. Thus, if the fundamental is $\cos x$, it is impossible to have a term of $\cos\left(\frac{3}{2}x\right)$. This makes computation easier since once the fundamental frequency is known all others become obvious. There is, however, a drawback in that more terms than might otherwise be necessary have to be used to describe adequately a particular function. For example, a function consists of $\cos t + 2 \cos 1.5t$. Since in Fourier series notation all terms have to be harmonics of some fundamental, the function would have to be expressed as $0 \cos (.5t) + 1 \cos 2(.5t) + 2 \cos 3(.5t)$. Thus, even though the result is the same, since the first term is zero, it is nevertheless necessary to deal with three rather than two terms.

Though there are disadvantages, such as in the example above and the Gibbs phenomenon for discontinuous functions discussed in Chapter 3, the Fourier series provides the closest approximation to an arbitrary functions $f(t)$.

least squared error

In actual practice we cannot deal with an infinite number of terms, so we must ask what will provide the best fit. It turns out that the Fourier series will give the least squared error approximation. A demonstration of this appears at the end of this chapter. Thus, even though in some instances this series has some disadvantages, it is the best general solution to practical problems.

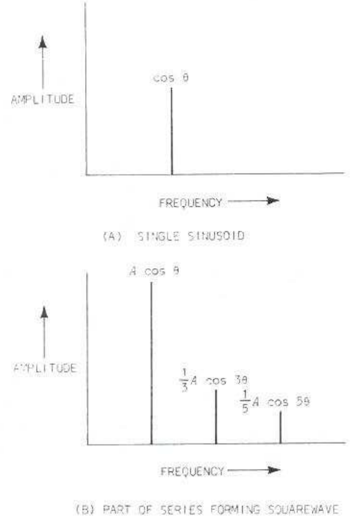


Fig. 2-9. Frequency-domain representations (phase not shown).

DO SPECTRAL LINES EXIST

According to Fourier theory, a squarewave consists of a fundamental sinewave, having the same period as the squarewave, and odd harmonics whose amplitudes decrease in proportion to harmonic number. Mathematically, a squarewave is said to consist of $\frac{2\,V_{t0}}{T}\left(\frac{1}{2}+\frac{2}{\pi}\cos\theta-\frac{2}{3\pi}\cos 3\theta\ \ldots\right)$ as discussed in more detail in the next chapter. The frequency-domain representation along with that of a single sinusoid, cos θ, is shown in Fig. 2-9. Few people have difficulty in visualizing the physical existence of the sinusoid represented in Fig. 2-9A. On the other hand, many of us have difficulty in visualizing the physical existence of the sinusoids,[7] or *spectral lines* as these are often called, represented in Fig. 2-9B. "I know," the statement often goes, "that a squarewave can be treated as if it were made up of sinewaves, but do the sinewaves *really* exist?" This is a difficult philosophical question, which fortunately need not be resolved for the practical utilization of Fourier techniques. This is because *linear time-invariant circuits*, which we are talking about, *behave as if these sinewaves did in fact exist.*

spectral lines

[7] R. W. Cushman, "Spectrum Analyzer Myths," Report No. NADC-EL-6452.

The fact that many time-variable circuits behave as if spectral lines did not exist is used as "proof" by those who choose not to believe in their existence. The believers, on the other hand, argue that all the above proves is that time-variable circuits are poor "detectors" of spectral lines. The fact that a man cannot detect light does not necessarily mean that there is no light. We get the same result if the man has poor detectors, that is — he's blind. What the question comes down to is the ancient one of *primary* and *secondary* qualities. The primary qualities are those that really exist in an object or phenomenon, while the secondary qualities are those which only seem to exist by virtue of interaction with the detector. This question has acquired modern importance in the area of quantum mechanics where the role of the observer is of the greatest significance.

The resolution of the question — "Does thunder make a sound when there is no one there to hear it?" — may have important philosophical implications, but it contributes very little to a practical discussion on spectrum analysis. The same may be said regarding the question on the *real* existence of spectral components. Rather, the question that should be asked is: "Do real circuits behave as if spectral components exist?" This is taken up in the next section.

RESPONSE OF CIRCUITS TO SIGNALS

linear time-invariant circuits

First it is necessary to emphasize that by circuits is meant *linear time-invariant* circuits. Fourier theory does not necessarily apply to nonlinear or time-variable circuits. Though all circuit elements eventually become nonlinear, and no physical resistor will obey Ohm's law at infinitely large voltages, it is fortunate that for the range of signals of interest a majority of circuits are linear. What is of interest then is how linear time-invariant circuits will behave under the stimulus of an arbitrary signal input.

steady state

transient

Generally a circuit which is excited by an arbitrary input will respond in two distinct ways. One is called the *steady-state* response, while the other is the *transient* response. Most people, when confronted by the words steady state and transient think of a long time interval and a short time interval respectively. Though it is true that the transient state is usually characterized by a short

time interval, this type of classification can be misleading since there can be conditions where a steady state is never reached. Of greater importance to our discussion, is the fact that transient behavior is determined by the so-called force-free solution; meaning essentially that the basic characteristics of the transient response are determined completely by the circuit parameters. Thus, if a current pulse is injected into an R, L, C circuit formed into a loop, damped oscillations at radian frequency $\omega = \frac{1}{\sqrt{LC}}$ will result when $\left(\frac{R}{2L}\right)^2 < \frac{1}{LC}$. The shape and amplitude of the current pulse will help determine the amplitude of the oscillations, but the oscillating frequency is independent of the forcing function. This is a very important point because it leads to the following results. In order to test for the existence of spectral lines, we perform an experiment in which a set of very narrowband contiguous filters is subjected to various inputs. What should be the result? Based on the previous discussion on the transient response, the output of each filter will be either nothing or a damped sinusoid at filter frequency; nothing else is possible. If the filter bandwidths are sufficiently narrow, the damping will be very slight and for all intents and purposes it will appear as if we are dealing with a continuous wave. Actually, the result could have been anticipated without any knowledge of transient behavior. Obviously, a filter can only have an output within its passband, that is what is meant by the word filter. But one can go further. Computing the output amplitude distribution as a function of the type of input, one finds the remarkable result that the transient response for an arbitrary input is identical to the steady-state response when the steady-state response is computed for an input composed of the Fourier components of the original arbitrary input[8]. Hence — *real, linear, time-invariant circuits behave as if spectral lines did in fact exist.*

[8] See Weber, *Linear Transient Analysis*, Vol II, for a discussion on the response of ideal filters to pulses.

If one chooses to believe in the real existence of spectral lines, then a Fourier analysis is simply a computation of some of the parameters of a signal. These parameters are eventually used in the practical business of determining the steady-state response of some network. If, on the other hand, one chooses not to believe in the real existence of spectral lines, then a Fourier analysis is simply a mathematical procedure that has no counterpart in physical reality. It is just a convenient technique for solving problems similar, for example, to the laying out of an orderly array of numbers in determinants when solving simultaneous equations. The solution of the equations may correspond to something physical, but it certainly is not necessary to validate the technique of solution by finding some entity which is physically spread out in an orderly array similar to the determinant.

The great utility of the Fourier technique is that it permits the solution of complicated transient problems by relatively simple steady-state techniques. This alone is sufficient justification for its use. If, in addition, one believes in the existence of spectral lines, then the advantages of Fourier techniques are obvious.

spectrum analyzer

All the arguments advanced for the use of Fourier techniques also hold true for the use of the instrument called a spectrum analyzer. Again one can look at this in two ways. One is that the display shows the spectral or energy distribution of the signal. The second view is that the display is the transient response of the spectrum analyzer circuits in response to the stimulus of the signal. Both views lead to the same final result: We can use the spectrum analyzer display to compute or predict the response of various linear time-invariant circuits under the same stimulus. This is the only justification necessary for the use of this instrument.

APPENDIX

1) ORTHOGONAL FUNCTIONS

The series $1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x \ldots$ is a series of orthogonal functions between the limits ± 1.

For any series of orthogonal functions:

$$\int_a^b f_m(x)\, f_n(x)\, dx = 0, \qquad m \neq n$$

For the above series we have:

$$\int_{-1}^{+1} (1)(x)\, dx = \left[\frac{1}{2}x^2\right]_{-1}^{+1} = \frac{1}{2} - \frac{1}{2} = 0$$

$$\int_{-1}^{+1} (1)\left(x^2 - \frac{1}{3}\right) dx = \left[\frac{1}{3}x^3 - \frac{1}{3}x\right]_{-1}^{+1} = 0$$

$$\int_{-1}^{+1} (x)\left(x^2 - \frac{1}{3}\right) dx = \left[\frac{1}{4}x^4 - \frac{1}{6}x^2\right]_{-1}^{+1} = 0$$

$$\int_{-1}^{+1} (1)(1)\, dx = \left[x\right]_{-1}^{+1} = 2$$

$$\int_{-1}^{+1} (x)(x)\, dx = \left[\frac{1}{3}x^3\right]_{-1}^{+1} = \frac{2}{3}$$

$$\int_{-1}^{+1} \left(x^2 - \frac{1}{3}\right)\left(x^2 - \frac{1}{3}\right) dx = \left[\frac{1}{5}x^5 - \frac{2}{9}x^3 + \frac{1}{9}x\right]_{-1}^{+1} = \frac{8}{45}$$

showing that the integral of the product of different terms is zero while the integral of a term times itself is not zero. Hence, the series is composed of a set of orthogonal functions.

2) COMPLEX NOTATION

The Euler identity

$$\epsilon^{j\theta} = \cos\theta + j\sin\theta$$

can be used to compute the values of various complex expressions. Thus:

$$\epsilon^{j\pi} = \cos\pi + j\sin\pi = -1 + j0 = -1$$

Similarly, the reader can verify that:

$$\epsilon^{j(\pi/2)} = +j$$

$$\epsilon^{-j(\pi/2)} = -j$$

$$\epsilon^{-j(3\pi/2)} = +j$$

$$\epsilon^{j2\pi} = +1$$

Besides the trigonometric form, complex quantities can also be expressed in algebraic form. Thus $Z = a + jb$ is a complex quantity, $Z^* = a - jb$ is called the *conjugate* of Z.

A complex quantity and its conjugate have a specific relationship to each other. For example:

$$Z + Z^* = (a + jb) + (a - jb) = 2a$$

Likewise, keeping in mind that $j = \sqrt{-1}$, the reader can verify that

$$Z \cdot Z^* = a^2 + b^2$$

$$Z - Z^* = j2b$$

3) PROOF THAT THE TRUNCATED FOURIER SERIES PROVIDES A LEAST SQUARED ERROR FIT

Let $f(t)$ be an arbitrary function of time which is to be expanded in an infinite series of sines and cosines. If we take only a finite number of terms, then the series is only approximately equal to $f(t)$. Thus:

$$f(t) \cong \sum_{m=0}^{n} [a_m \cos m\,\omega t + b_m \sin m\,\omega t],$$

where the terms being summed is an arbitrary collection of sines and cosines, and not necessarily a Fourier series. We wish to show that in order to have the least squared error between $f(t)$ and the noninfinite series, which stops at $m = n$, approximating $f(t)$, it is necessary that the series be a Fourier series. Thus, the squared error is

$$\rho = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} [f(t) - \Sigma]^2\, dt$$

where the sigma (Σ) is a shorthand notation for the truncated series. We wish the error (ρ) to be minimum, which means that the differential of the error should be zero. Thus, finding the minimum error with respect to the coefficients of the cosine terms (a_m)

$$\frac{\partial \rho}{\partial a_m} = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} \frac{\partial}{\partial a_m} [f(t) - \Sigma]^2\, dt$$

$$= \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} 2[f(t) - \Sigma](-\cos m\,\omega t)dt = 0$$

Separating the terms, we have

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} f(t) \cos m\,\omega t\, dt = \int_{-\frac{T}{2}}^{+\frac{T}{2}} (\Sigma) \cos m\,\omega t\, dt$$

The summation sign, of course, stands for

$$\sum_{m=0}^{n} [a_m \cos m\,\omega t + b_m \sin m\,\omega t]$$

Now the definite integral, taken over one period, of a sine-cosine product is zero. Likewise, by virtue of orthogonality, are all cosine-cosine products except those where the two terms are the same. Hence

$$\int_{-\frac{T}{2}}^{+\frac{T}{2}} f(t) \cos m\,\omega t\, dt = \int_{-\frac{T}{2}}^{+\frac{T}{2}} a_m \cos^2 m\,\omega t\, dt = \frac{a_m T}{2}.$$

Solving for a_m, we have

$$a_m = \frac{1}{\frac{T}{2}} \int_{-\frac{T}{2}}^{+\frac{T}{2}} f(t) \cos m\,\omega t\, dt.$$

Likewise, it can be shown that

$$b_m = \frac{1}{\frac{T}{2}} \int_{-\frac{T}{2}}^{+\frac{T}{2}} f(t) \sin m\,\omega t\, dt,$$

which are the same as the Fourier coefficients, hence the Fourier series gives the best fit[9].

---

[9] A proof showing indentity between individual terms will be found in Weber, *Linear Transient Analysis*, Vol. I, pages 258-259.

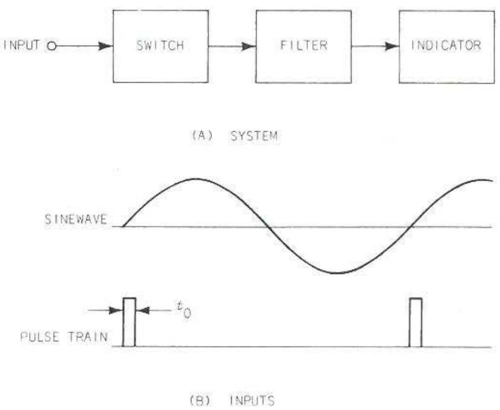


Fig. 2-10. Time-variable network.

4) EXAMPLE OF TIME-VARIABLE NETWORK

Consider the system illustrated in Fig. 2-10. Suppose the input is a single continuous sinewave at a frequency within the passband of the filter and the switch is closed. Obviously, the indicator shows a response. Now we leave the switch open for a short period of time during each cycle. The indicator may read something different than before, but all will agree that the indicator continues to show an output.

Now let the input be changed to a train of narrow pulses having the same period as the sinewave. According to Fourier theory, the pulse train can be considered as the sum of an infinite number of sinewaves. One of these Fourier sinewaves, which has the special name of fundamental, should be at the same frequency as the original sinewave since the pulse train has the same period as the original sinewave.

The switch is closed and we look at the indicator. Sure enough, there is a response, which means that the fundamental is there. Now we leave the switch open for a short period of time (t_0) during each cycle, so arranged that the switch opening coincides with the occurrence of a pulse. Unlike the case of the single sinewave, the indicator now shows nothing. Obviously, we don't get the same results for both inputs. The reason for this is that the switch makes this a time-variable network for which Fourier theory does not apply.

3

FOURIER ANALYSIS

All practical functions defined over an interval, such as $-\pi$ to $+\pi$, can be expanded in a Fourier series. This holds true even when the waveform represented by the function is nonrepetitive and exists only during the defined interval. The interpretation of the Fourier series of these isolated pulses has no meaning outside the defined interval. Single pulses can also be treated from the continuous spectrum Fourier integral point of view, and this is the approach that will be taken here. Thus, Fourier series will be considered for repetitive functions only.

FOURIER SERIES

Practical, physically realizable, functions having a period 2π can be expanded in a series of trigonometric functions such that the function

$$f(x) = \frac{a_0}{2} + (a_1 \cos x + b_1 \sin x)$$

$$+ (a_2 \cos 2x + b_2 \sin 2x)$$

$$\ldots + (a_n \cos nx + b_n \sin nx), \qquad (3\text{-}1)$$

which can be represented in summation form

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

where x is an angle in radians.

The series defined by equation (3-1) is known as a *Fourier series.* The coefficients, a_n and b_n, are constants which are determined by the form of the original function $f(x)$. The two summations, sine and cosine, can be combined into a single series by the addition of a phase angle. Thus:

$$f(x) = \frac{C_0}{2} + \sum_{n=1}^{\infty} C_n \cos (nx + \phi_n), \tag{3-2}$$

where:

$$C_0 = a_0,$$

$$C_n = \sqrt{a_n^2 + b_n^2},$$

$$\phi_n = \tan^{-1} \left(\frac{-b_n}{a_n}\right).$$

phase angle

Expression (3-2) is the more useful for spectrum analyzer work since the spectrum analyzer displays the combined amplitude, C_n, rather than the separate sine and cosine amplitudes. Furthermore, spectrum analyzers of the type under discussion do not display any phase characteristics. Therefore, except in special cases such as the combination of several complicated spectra, the phase angle, ϕ, will be ignored.

As is demonstrated at the end of this chapter, the coefficients a_n and b_n are related to the original function $f(x)$ through the following integrals.

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\cos nx\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos nx\,dx \tag{3-3}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\sin nx\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin nx\,dx$$

period

As indicated in (3-3) above, the integration can be carried out between $+\pi$ and $-\pi$ or between 0 and 2π. The choice depends on how the original function $f(x)$, is defined and on which integration involves less work. In any event, the final result is the same, regardless of which integration is used. Very often the period is some arbitrary time interval T rather than 2π. The Fourier series still applies except that all expressions have to be scaled by the factor $2\pi/T$, thus:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos 2n\frac{\pi}{T}x + b_n \sin 2n\frac{\pi}{T}\right) \tag{3-4}$$

and the coefficients are given by

$$a_n = \frac{2}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} f(x)\cos\frac{2n\pi x}{T}\,dx = \frac{2}{T}\int_0^{T} f(x)\cos\frac{2n\pi x}{T}\,dx, \tag{3-5}$$

$$b_n = \frac{2}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} f(x)\sin\frac{2n\pi x}{T}\,dx = \frac{2}{T}\int_0^{T} f(x)\sin\frac{2n\pi x}{T}\,dx.$$

Again a_n and b_n can be combined into a single amplitude factor $C_n = \sqrt{a_n^2 + b_n^2}$, which is the factor displayed on the spectrum analyzer.

When using the C_n representation it is still necessary to solve two separate equations, one for the a_n terms and the other for the b_n terms. These two equations can be combined into one by the use of complex notation.

Thus in complex notation:

$$f(x) = \sum_{n=-\infty}^{n=+\infty} d_n \epsilon^{jnx}, \tag{3-6}$$

where the coefficient d_n is obtained from

$$d_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x)\epsilon^{-jnx}\, dx \tag{3-7}$$

and n takes on all positive and negative values as well as zero. Thus, except for the case where $n = 0$, there are still two terms for every n, namely d_n and d_{-n} with one being the conjugate of the other. The two complex coefficients, when combined in accordance with the rules for complex numbers (as discussed in Chapter 2), leads back to the more familiar trigonometric expression

$$\begin{aligned} d_n\epsilon^{jnx} + d_{-n}\epsilon^{-jnx} &= \frac{a_n - jb_n}{2}\epsilon^{jnx} + \frac{a_n + jb_n}{2}\epsilon^{-jnx} \\ &= a_n \cos nx + b_n \sin nx. \end{aligned} \tag{3-8}$$

The complex notation presents some conceptual difficulty because of the appearance of what seems to be negative frequencies, since the summation goes over negative as well as positive numbers. Some ideas on the interpretation of negative n's will be found in Chapter 2. On the other hand, the complex expressions (3-6) and (3-7) are much more compact than their trigonometric counterpart. The complex notation is particularly useful in the Fourier integral representation which is needed for the analysis of continuous spectra.

FOURIER APPLICATIONS

sinusoid

In a Fourier series representation, the fundamental waveform is the sinusoid. Hence, it is of interest to determine the frequency-domain representation, by way of Fourier series, for a sinusoid. The time-domain function is:

$$f(t) = A \cos \omega_0 t$$

The frequency-domain function is the Fourier series previously defined. Using the trigonometric representation, equation (3-5) is used to determine the Fourier coefficients a_n and b_n. Equation (3-5) reproduced here as equation (3-10) is:

$$a_0 = \frac{2}{T}\int_0^{+T} f(x)\, dx \tag{3-9}$$

$$\begin{aligned} a_n &= \frac{2}{T}\int_0^{+T} f(x) \cos \frac{2n\pi x}{T}\, dx \\ b_n &= \frac{2}{T}\int_0^{+T} f(x) \sin \frac{2n\pi x}{T}\, dx. \end{aligned} \tag{3-10}$$

The coefficient a_0 is just the average value of the function $f(x)$. In engineering language we could say that a_0 is the DC level of the waveform. Thus by inspection, $a_0=0$ since the average value of a sinusoid is zero. Just as the solution for a_0 is simple, so is the solution for the other coefficients difficult. This is because we are dealing not with a cosine pulse but a continuous cosine wave. In engineering language, this is called a CW signal and is assumed to exist forever. This leads to integration of sinusoids with infinite limits which cannot be performed directly. Actually, the solution is easier to obtain using the complex notation; equation (3-10) was only used to show the simple nature of a_0. In any event, without going through the complicated solution, we shall simply state the result: The frequency-domain representation for a sinusoid is an impulse function:

$$f(t) = A \cos \omega_0 t$$

$$\omega_0 = 2\pi f_0 t$$

$$F(\omega) = \frac{A}{2}\left[\delta(f + f_0) + \delta(f - f_0)\right] \tag{3-11}$$

impulse function

The function $\delta(x)$ is called a Dirac, delta, or *impulse function*, where $\delta(f)$ is a frequency impulse and $\delta(t)$ is a time impulse. The reason for the two parts in the Fourier representation is that in the solution of problems by the complex notation method we get two coefficients, these are d_n and d_{-n} or in this case d_{+f0} and d_{-f0}. Since in practical applications one is interested in the composite amplitude, one might as well use the notation $A\ \delta(f - f_0)$, meaning an impulse at frequency $f = f_0$ and of strength A. An impulse has a finite area, in this case equal to A, and zero width, thus requiring infinite amplitude. Obviously, the impulse is only a theoretical function since infinite amplitudes cannot be generated. This does not mean that the impulse does not have validity. The impulse function is just as valid for analysis and theoretical representation purposes as the sinusoid of infinite time duration whose frequency representation it is. Certainly infinite-duration sinusoids do not exist in practice, but that doesn't prevent using these to represent practical signals. Some properties of impulse functions will be found in the appendix for this chapter.

The graphical representation of the impulse is symbolic rather than exact. This is because there is no way to faithfully graph a function that calls for infinite amplitude. A symbolic rather than exact graphical representation may disturb some people, but actually this is often done. The infinite-duration sinewave, for example, is represented symbolically by a finite number of cycles rather than the infinite number of cycles that it is supposed to have. In any event, what is of interest in the case of the impulse is not its amplitude, which is always considered infinite, but rather the *area* or *strength*, which is representative of the amplitude of its time-domain sinewave equivalent. Fig. 3-1 shows the time and frequency-domain representations of a simple CW signal, namely the sinewave and impulse. Fig. 3-1C represents a spectrum analyzer display. Note the identity between the Fourier series derived Fig. 3-1C and the desired representation shown in Fig. 2-9A.

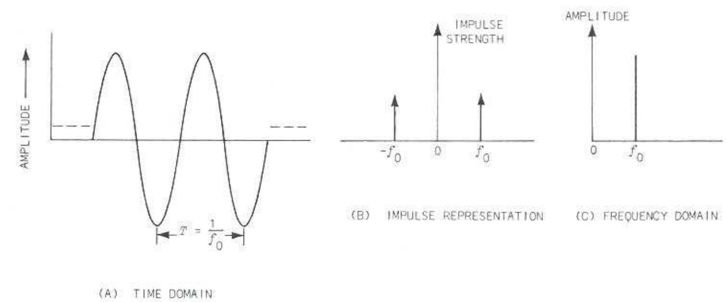


Fig. 3-1. Time- and frequency-domain representations for sinusoid.

rectangular pulse train

Let us consider now the important case of a rectangular pulse train of arbitrary pulse width (t_0) and arbitrary period (T) as shown in Fig. 3-2A.

The function $f(x)$ is defined by the amplitude A for

$-\frac{t_0}{2} < t < +\frac{t_0}{2}$ and zero everwhere else.

Using the scale factor $\frac{2\pi}{T}$ in equation (3-7), we have

$$d_n = \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} f(t)\,\epsilon^{-jn\frac{2\pi t}{T}}\,dt \qquad (3\text{-}12)$$

which is the complex notation equivalent of equation (3-5).

Since the pulse only exists between the limits of $\pm\frac{t_0}{2}$ we only need to integrate between these limits, thus:

$$d_n = \frac{1}{T}\int_{-\frac{t_0}{2}}^{+\frac{t_0}{2}} A\,\epsilon^{-jn\frac{2\pi t}{T}}\,dt = \frac{A}{T}\,\frac{1}{-jn\frac{2\pi}{T}}\,\epsilon^{-jn\frac{2\pi t}{T}}\Bigg]_{-\frac{t_0}{2}}^{+\frac{t_0}{2}}$$

substituting the limits:

$$d_n = \frac{A}{T}\left(\frac{\epsilon^{-jn\frac{2\pi}{T}\frac{t_0}{2}} - \epsilon^{+jn\frac{2\pi}{T}\frac{t_0}{2}}}{-jn\frac{2\pi}{T}}\right)$$

Rearranging terms and bringing the minus sign from the denominator to the numerator:

$$d_n = \frac{A}{\frac{n\pi T}{T}}\left(\frac{\epsilon^{+jn\frac{2\pi}{T}\frac{t_0}{2}} - \epsilon^{-jn\frac{2\pi}{T}\frac{t_0}{2}}}{2j}\right)$$

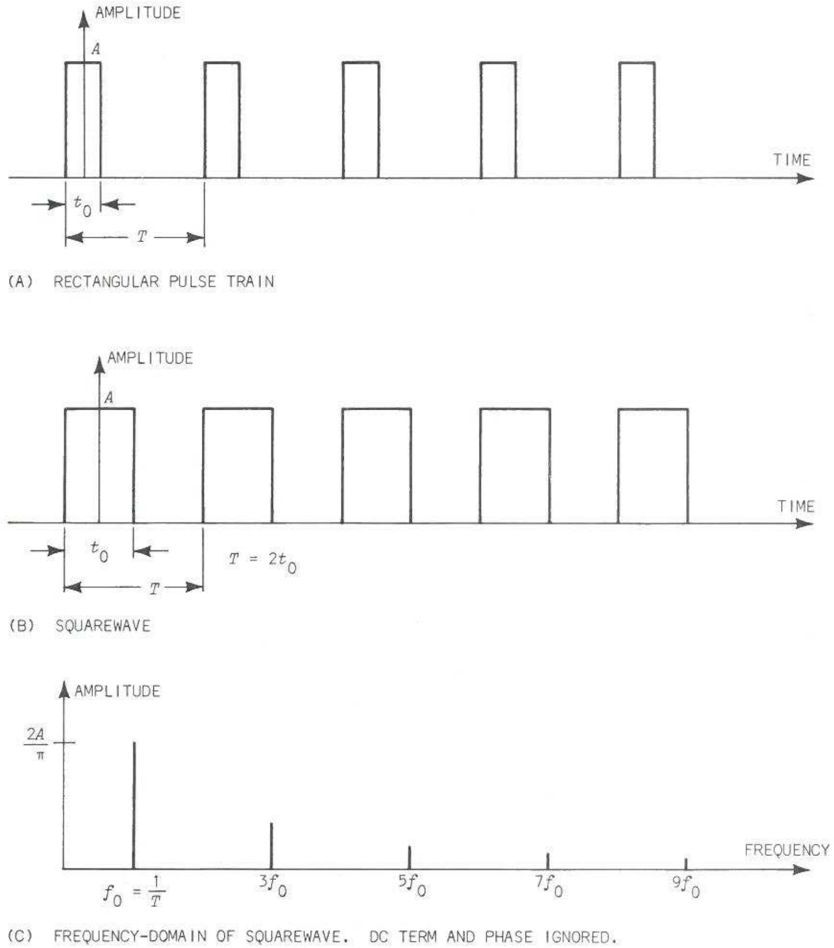


Fig. 3-2. Time- and frequency-domain representations of a pulse train.

The part in the brackets is the sine function as discussed in the appendix to Chapter 2. Hence:

$$d_n = \frac{A}{n\pi} \sin \frac{n\pi t_0}{T}$$

The same result is obtained for the d_{-n} components.

What is of interest is the overall amplitude term C_n, which is obtained by summing d_n and d_{-n}. Thus after adding and rearranging terms, the final result is:

$$C_n = \frac{2At_0}{T} \frac{\sin \frac{n\pi t_0}{T}}{\frac{n\pi t_0}{T}} \tag{3-13}$$

This is the origin of the so-called sine x over x $\left(\frac{\sin x}{x}\right)$ distribution which will be used extensively in connection with continuous spectra.

interpreting results

The interpretation of (3-13) is that the pulse train is made up of a DC component C_0 and a set of sinusoids with amplitudes C_1, $C_2 \cdots C_n$. The term C_0 is usually ignored in graphical representations, certainly it does not show up on the spectrum analyzer. Therefore, a graph of a frequency distribution such as (3-13) would consist of a representation of a train of sinusoids. A sinusoid, as indicated in the previous example, is represented in the frequency domain by an impulse. However, it is the area rather than the amplitude of the impulse which is equivalent to the amplitude of the sinusoid.

A graphical display of a Fourier series, such as (3-13), therefore consists of a set of vertical lines which are symbolic impulses. One impulse is used per sinusoid in the series. As an example, let $T = 2t_0$, making the pulse train into a squarewave as shown in Fig. 3-2B. Substituting $T = 2t_0$ into (3-13) we get:

$$C_n = A \frac{\sin \frac{n\pi}{2}}{\frac{n\pi}{2}} \tag{3-14}$$

as the equation for the amplitudes of the sinusoids which when combined make up the squarewave. The individual amplitudes are obtained by substituting for n. Thus, for the fundamental, $n = 1$, we get

$$C_1 = A \frac{\sin \frac{\pi}{2}}{\frac{\pi}{2}} = \frac{2A}{\pi} \quad \left(\text{since } \sin \frac{\pi}{2} = 1\right).$$

For the second harmonic, amplitude $n = 2$ is substituted, leading to:

$$C_2 = A \frac{\sin \pi}{\pi} = 0 \quad (\text{since } \sin \pi = 0).$$

When a similar procedure is followed for the other harmonics, it is observed that all the even harmonics are zero, while the amplitude of the odd harmonics progresses as $1/n$, so that the fifth harmonic is one-fifth as large as the fundamental while the ninth harmonic amplitude is one ninth as large as the fundamental, etc. In addition, there is also the DC or average term at $n = 0$. This term is difficult to obtain directly from equation (3-14), since substitution of $n = 0$ leads to the indeterminant zero over zero. However, this term is easily obtained by observation of Fig. 3-2B, it is simply t_0/T or $1/2$ for a squarewave.

The complete frequency-domain representation for a squarewave, therefore, is:

$$f(t) = A\left(\frac{1}{2} + \frac{2}{\pi} \cos \omega_0 t - \frac{2}{3\pi} \cos 3\omega_0 t \cdots\right) \tag{3-15}$$

where

$$\omega_0 = \frac{2\pi}{T}$$

It should be kept in mind that although (3-15) shows a DC term, the spectrum analyzer will not show this. Although (3-14) shows alternating 180° phase reversals as indicated by the alternating plus and minus signs, the spectrum analyzer will not show this either. Fig. 3-2C shows the frequency-domain characteristics of the squarewave.

super-
position

One of the most important concepts that is applicable to Fourier analysis is that of superposition. *Superposition* essentially means the simultaneous existence of signals, where the combined effect or composite signal is obtained by the addition of the several components. When the individual components are expressible mathematically by ordinary algebraic functions, simple addition is all that is needed. When the signal components are vectors represented by complex notation, then the rules for vectorial addition, which take into account phase relationships, must be used.

Superposition is actually tacitly assumed in the formulation of Fourier series. This is because the statement that a complex waveform can be represented by a sum of sinewaves cannot be made unless superposition holds so that the sinewaves can be added. A more significant use of superposition is in the relationship of cause and effect. Thus, in using Fourier series in the solution of network response problems, the solution is frequently obtained by performing a Fourier analysis of the input signal and then summing the network responses obtained by considering each of the Fourier sinewave components as being applied to the network singly. Of more importance in spectrum analysis is the fact that the spectra of complex waveforms can be obtained by superposition. Thus, if the Fourier representation for a time-domain function $f(t)$ is $F(f)$ and for another time-domain function $g(t)$ it is $G(f)$, then the Fourier representation for the combined time-domain function $f(t) + g(t)$ is just $F(f) + G(f)$. This, when combined with the fact that a time delay introduces a phase shift but otherwise leaves the spectrum unaffected, permits the computation of complex spectra simply by the addition of simpler spectra. Thus, for example, the frequency-domain characteristics of a trapezoid can be found by the addition of the spectra of two triangular pulses and one rectangular pulse.

As a specific example, consider a squarewave as constructed by the addition of two sawtooth waves shown in Fig. 3-3. This squarewave is identical to that of Fig. 3-2B except for the elimination of the DC term. The Fourier coefficients for a sawtooth, which can be obtained by the standard method of integration, are given by:

$$C_n = A\frac{1}{\pi n}$$

where A is the maximum height (shown as 2 in Fig. 3-3).

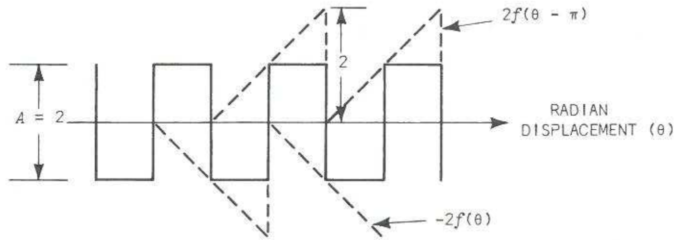


Fig. 3-3. Squarewave as sum of two sawtooth waves.

The sum of the two Fourier series is:

$$f(\theta) = \frac{2}{\pi}\left[\frac{1}{2} - \sin(\theta - \pi) - \frac{1}{2}\sin 2(\theta - \pi) - \frac{1}{3}\sin 3(\theta - \pi)\cdots\right]$$
$$- \frac{2}{\pi}\left(\frac{1}{2} - \sin\theta - \frac{1}{2}\sin 2\theta - \frac{1}{3}\sin 3\theta\cdots\right) \qquad (3\text{-}16)$$

The DC terms of the sawtooth waves are equal and opposite and hence cancel. The bottom sawtooth has a half-cycle phase shift with respect to the top one, hence the $(\theta - \pi)$ term.

The two sinewave series in (3-16) can be combined into a single series with the help of the trigonometric identity:

$$\sin n(x - \pi) = (-1)^n \sin(nx) \qquad (3\text{-}17)$$

Using (3-17), equation (3-16) becomes:

$$f(\theta) = \frac{2}{\pi}\left(\frac{\sin\theta}{1} + \frac{\sin 2\theta}{2} + \frac{\sin 3\theta}{3}\cdots\right)$$
$$- \frac{2}{\pi}\left(-\frac{\sin\theta}{1} + \frac{\sin 2\theta}{2} - \frac{\sin 3\theta}{3}\cdots\right). \qquad (3\text{-}18)$$

The series in (3-18) can be added term by term resulting in

$$f(\theta) = \frac{4}{\pi}\left(\frac{\sin\theta}{1} + \frac{\sin 3\theta}{3} + \frac{\sin 5\theta}{5}\right). \qquad (3\text{-}19)$$

Except for the loss of the DC term because of the vertical shift of the squarewave, the use of θ instead of ωt to simplify the notation, and the phase shift resulting from considering zero time at the start rather than in the middle of a cycle as in Fig. 3-2, equations (3-15) and (3-19) are the same. Actually, the analysis could have easily been arranged so that (3-15) and (3-19) would be identical.

However, besides illustrating the use of the principle of superposition, it is the intent to illustrate that regardless of how the analysis is performed the essential features, which are those that are displayed on a spectrum analyzer, remain the same. The spectrum analyzer does not display the DC term, nor is the spectrum analyzer sensitive to phase, so that the switch from cosine to sine has no effect. The important part of the analysis is that a squarewave is represented by an infinite series of sinusoids consisting of a fundamental and odd harmonics, with harmonic amplitude decreasing as the inverse of the harmonic number. This is precisely the information that can be obtained by means of the spectrum analyzer.

Gibbs phenomenon

Based on the previous discussion, one gets the impression that one should be able to reconstruct any waveform simply by adding the sinusoids forming the Fourier series. Of course, since most Fourier series call for an infinite number of terms, such a reconstruction is not practical.

Nevertheless, this can be considered theoretically. Such a study helps in establishing the validity of the Fourier approach and is useful in establishing guidelines on how many terms of the series can be considered a sufficiently close approximation. When this is done for a discontinuous function, such as a squarewave, the result is not Fig. 3-2B but rather that shown in Fig. 3-4.

As the number of harmonics in the summation is increased the resultant waveform is seen to oscillate around the discontinuities at the corners as shown in Fig. 3-4A. This occurs because in the vicinity of a finite discontinuity the sum of the Fourier terms converge to the average value as the discontinuity is approached from both sides. As the number of terms in the series is increased the oscillations squeeze closer and closer together, and as the number of terms approaches infinity the oscillation is squeezed into a straight line as shown in Fig. 3-4B. This oscillating overshoot phenomenon is known as the *Gibbs phenomenon* in Fourier series. There is no way to get away from this overshoot when summing a Fourier series, except by modifying the coefficients; in which case it is of course no longer a straightforward Fourier series[1].

[1] A discussion on this and other aspects of the Gibbs phenomenon will be found in Guillemin's *The Mathematics of Circuit Analysis.*

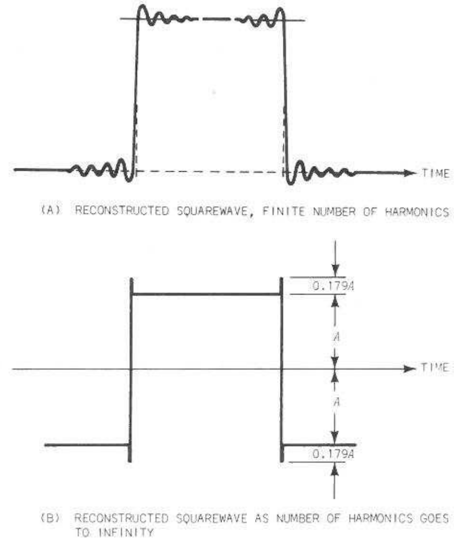


Fig. 3-4. Squarewave reconstructed from Fourier components.

The size of the overshoot, as the limit in the number of terms in the series is increased without bound, has been computed by many people including Weber, *Linear Transient Analysis*, vol. I. Under the best circumstances, the overshoot is about 18% as shown in Fig. 3-4B.

The fact that the sum of the Fourier terms does not seem to lead back to the original function appears, at first glance, to be a serious blow to Fourier theory. Actually from an energy point of view there is no discrepancy. This is because the overshoot is an infinitely thin line of finite amplitude and, hence, has zero area and no energy. It can of course be argued that infinitely thin lines do not exist in practice, but then neither do the infinitely steep slopes of perfect squarewaves. What it eventually comes down to is the question of how real physically realizable circuits behave, which is discussed in Chapter 2. It might be well to repeat from there the major point on which all spectrum analyzer work is based: *Physically realizable, linear, time-invariant networks behave as if Fourier spectral components exist. The function of the spectrum analyzer is to provide information on the behavior of such circuits. Therefore, it is not necessary to believe in the "real" existence of Fourier spectral lines in order to accept the validity of the spectrum analyzer display.*

continuous dense spectrum

While the concept of a frequency spectrum for a pulse train is at least intuitively acceptable, the concept of a frequency spectrum for a single pulse is much more difficult to comprehend. Indeed the frequency distributions cannot be treated in the same manner, since the former is a discrete or line-type spectrum while the latter is a continuous dense type of spectrum. The mathematical treatment for these two types of spectra is also different, since the Fourier series is not directly applicable to the continous spectrum.

The continuous spectrum is handled easiest when considered as the limiting case of a discrete spectrum. Let us, therefore, start with the frequency-domain representation of a pulse train such as shown in Fig. 3-2A. The Fourier coefficients are given by equation (3-13), which is reproduced below:

$$C_n = \frac{2At_0}{T} \frac{\sin \frac{n\pi t_0}{T}}{\frac{n\pi t_0}{T}} \tag{3-20}$$

Disregarding the DC term, which will not be displayed on the spectrum analyzer anyway, one obtains the amplitude of the fundamental and the various harmonics by substituting $n = 1, 2, 3 \cdots$ into (3-20). Thus, the amplitude of the fundamental is:

$$C_1 = \frac{2At_0}{T} \frac{\sin \frac{\pi t_0}{T}}{\frac{\pi t_0}{T}}. \tag{3-21}$$

The second harmonic amplitude is:

$$C_2 = \frac{2At_0}{T} \frac{\sin \frac{2\pi t_0}{T}}{\frac{2\pi t_0}{T}}$$

narrow pulses

and so forth. The important and interesting case occurs when t_0/T is small; in other words, when we are dealing with a train of narrow pulses rather than squarewaves. Under such conditions the angle $(\pi t_0/T)$ is small and (3-21) reduces to

$$C_1 = \frac{2At_0}{T} \tag{3-22}$$

since the sine of a small angle is essentially equal to the angle. This approximation only holds true in the vicinity of the fundamental, obviously the angle $(n\pi t_0/T)$ eventually gets large as the harmonic number (n) increases. As the harmonic number is increased, the quantity in the numerator of (3-20) decreases while the denominator increases, so that the amplitude C_n decreases. Eventually, a point is reached where $n = T/t_0$ so that the angle becomes $n\pi t_0/T = \pi$. Since $\sin \pi = 0$, the amplitude of that particular harmonic is zero. This is called a *spectrum null*, or simply a null. As n is increased further, the harmonic amplitude increases, goes through a peak, and decreases until at $n = 2T/t_0$, when the angle is equal to 2π, there is again a spectrum null. This process of peaks of decreasing amplitude and nulls continues ad infinitum. A plot of (3-20) for $t_0/T = 0.1$ is given in Fig. 3-5A. Just as for the spectrum of the squarewave train shown in Fig. 3-2C, each of the vertical lines in Fig. 3-5A represents the amplitude of a sinusoid.

spectrum null

Except for the Gibbs phenomenon, we get back the original pulse train when these sinusoids are added in appropriate phase. It should again be emphasized that the spectrum analyzer does not show phase, so phase information was not shown in Fig. 3-5. As expected, the 10th, 20th, 30th $\cdots$ harmonics, corresponding to an angle equal to multiples of π, go to zero. If the ratio of pulse width to interpulse period (t_0/T) were other than ten to one, other harmonics than numbers 10 or 20 would go to zero, but this would still occur when the angle is equal to a multiple of π. Thus, the total angle $(n\pi t_0/T)$ rather than the harmonic number (n) is the fundamental parameter. Therefore, as we contemplate the effect on the spectrum of changes in the ratio t_0/T, the horizontal scale will remain in units of total angle, representing radian frequency rather than harmonic number.

T/t_0 ratio

Consider now the effect on equation (3-20) of an increase in the interpulse spacing T. Suppose, for example, T is tripled so that $t_0/T = 1/30$ rather than $1/10$. Since all the harmonics are multiplied by the factor t_0/T, all amplitudes will decrease to one-third their original size. The basic shape of the spectral distribution will remain the same $(\sin x)/x$ shape. Nulls will again appear where x is a multiple of π, which happens when the harmonic number (n) is a multiple of T/t_0. As the ratio T/t_0 is increased, the number of components between nulls will also increase. In the above example, T/t_0 is increased to 30, so there are now thirty rather than ten harmonics between nulls. Or, looked at in another way, there are now three times as many signal components per unit frequency as there were before. Fig. 3-5B is a plot of the spectrum of a rectangular pulse train with $T/t_0 = 30$.

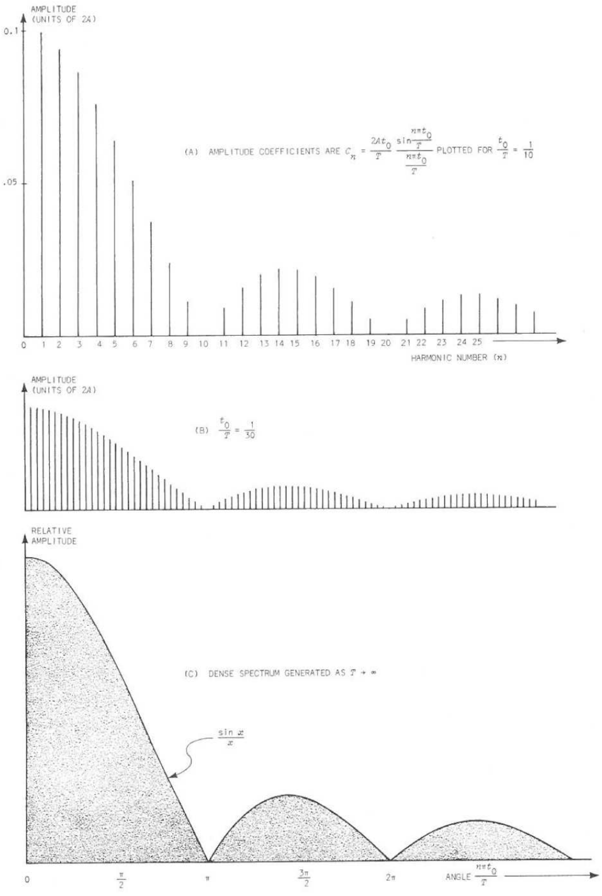


Fig. 3-5. Frequency distribution for rectangular pulse train.

As one proceeds to increase the ratio T/t_0 by increasing the interpulse spacing (T), three important things happen. As $T \to \infty$, these are: 1) the amplitude of the individual components approaches zero; 2) the number of harmonics between nulls increases without bound; and 3) the shape of the curve is unchanged, remaining the same $(\sin x)/x$.

Ignoring, for the moment, the apparent disappearance of everything as the amplitudes go to zero, let us concentrate on the meaning of the last two conclusions. As the number of harmonics increases without bound, a state is reached where it becomes impossible to distinguish between individual harmonics. Also, it would not make much sense to talk about individual harmonics as there are essentially an infinite number of these in any frequency interval, no matter how small. Yet, one parameter has remained unchanged as T was increased, this is the curve generated by the locus of the end points of the harmonic amplitudes. The shape of the curve is the previously discussed $(\sin x)/x$. Obviously the meaning of $x = n\pi t_0/T$ as $T \to \infty$ will have to be reinterpreted, otherwise x appears to approach zero. Ignoring this problem for the moment, we end with the graph of Fig. 3-5C.

dense spectrum

Fig. 3-5C is the frequency-domain representation of a *dense, continuous* spectrum. The spectrum is dense and continuous in the sense that, excepting the null points, no frequency can be found where there is no energy. Contrast this with the spectra of Figs. 3-5A and 3-5B where there is energy only at specified frequencies, as indicated by the harmonics, and zero everywhere else. The usual procedure in establishing a description of dense spectra is to first establish the mathematical validity of the Fourier integral equations which are then used in analyzing the spectra of single pulses. The procedure followed here is the opposite. First, using physical reasoning, we will develop an interpretation of the dense spectrum which can then be used to justify the use of Fourier integrals. This method, while not mathematically rigorous, is helpful in establishing how a spectrum analyzer works. Those interested in a rigorous derivation are referred to the references.

There are two basic points that need to be considered when establishing a physical interpretation of Fig. 3-5C: How to handle the angle $n\pi t_0/T$ and what to do with the apparent disappearance of the spectrum since the ratio t_0/T seems to approach zero as T gets infinitely large. First the matter of the angle. It should be recognized that as the interpulse spacing (T) increases, so does the number of harmonics (n) occurring over any arbitrary frequency range. For example, the number of harmonics between two null points, which occur at angular differences of π, is $n = T/t_0$. Thus, as T goes toward infinity so does n and the ratio (n/T) remains constant.

In order to get rid of the bothersome infinities it is, therefore, only necessary to treat the ratio (n/T) as a unit. This unit has the dimensions of inverse time interval which is frequency, hence n/T is designated by the symbol f. It should be recognized that the use of the symbol f has greater significance than simple dimensional correctness. The frequency (f) is actually the frequency at which the harmonic (n) occurs. For example, suppose $T = 1$ ms, then the fundamental is at $1/T = 1$ kHz, the second harmonic is at 2 kHz, the tenth harmonic at 10 kHz, etc., with the frequency of the nth harmonic at n/T. If, in addition, t_0 happened to be one-tenth the size of T, or 100 μs, then the tenth harmonic at 10 kHz would have zero amplitude according to Fig. 3-5A. Note that the frequency of the tenth harmonic is not affected by t_0, only n/T has to do with frequency, while t_0 determines amplitude. Based on the above reasoning, the angle $n\pi t_0/T$ is replaced by $\pi f t_0$, thus eliminating all problems with infinite T.

The amplitude coefficient of (3-20) is $2At_0/T$. Since A is pulse amplitude and t_0 is pulse width, the product At_0 is pulse area. The division by the interpulse period T is an averaging process, so that what is involved is the average pulse area. The factor 2 arises because theoretically the spectrum is symmetrical about the main lobe with center at $n = 0$. This point was discussed before in connection with the complex form of the Fourier series where for every coefficient of positive frequency d_n there is a corresponding conjugate of d_{-n}. Since, in practical spectrum analysis, negative frequencies have no meaning, the factors d_n and d_{-n} were combined to avoid confusion. For the rectangular pulse train d_{-n} and d_n are equal, which leads to an overall amplitude $C_n = 2d_n$. In any event the conceptual difficulty is not with the factor 2 but with $1/T$. The reason everything seems to go to zero is that equation (3-20) deals with the amplitude of individual harmonics, but, as previously discussed, individual harmonics have no meaning when dealing with a continuous spectrum. This is because there are apparently an infinite number of these. Obviously, if individual harmonics contained a finite amount of energy, no matter how small, the total energy of all the harmonics would become infinite, and that is physically impossible. Instead of dealing with individual harmonics it is necessary to deal with a *spectral density* or energy per unit bandwidth. This problem is analogous to that of the impulse function where the parameters are essentially zero width, infinite amplitude and finite area. Similarly, here the parameters are essentially zero amplitude, an infinite number of harmonics and finite total energy. When using the spectral-density concept, the bothersome $1/T$, actually frequency, is set equal to unity to represent a per-unit-bandwidth operation.

spectral density

Based on the above, the discrete spectral components of (3-20) are transformed to the continuous spectral-density distribution given below:

$$F(\omega) = 2At_0 \frac{\sin(\pi t_0 f)}{\pi t_0 f} \tag{3-22}$$

where $F(\omega)$ stands for a Fourier integral representation.

Fourier integral

The relationships for the complex form of the Fourier series are given in equations (3-6) and (3-7) and reproduced below:

$$f(x) = \sum_{n=-\infty}^{n=+\infty} d_n \, \epsilon^{jnx} \tag{3-23}$$

$$d_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x) \, \epsilon^{-jnx} \, dx \tag{3-24}$$

These equations are applicable when dealing with a discrete spectrum generated by a waveform having a finite period. This permits equation (3-23) to be the sum of a discrete series of sinusoids, one for each n as n takes on all the positive and negative integer numbers. The series is infinite in that there are an infinite number of integers, but the spectrum is not continuous since all except very specific values of n are forbidden.

The finite period is clearly evident from (3-24) where the limits of integration are $+\pi$ and $-\pi$. Here the period is assumed to be not only finite, but specifically 2π. The equation can, of course, be modified for an arbitrary period T rather than 2π, as shown in equation (3-12).

Fourier transform pair

These equations have to be modified when dealing with isolated pulses, whose period is essentially infinite and which have continuous rather than discrete spectra. The continuous nature of the spectrum requires integration rather than summation, while the limits have to be extended to include a time function which never seems to end. The two integral equations replacing (3-23) and (3-24) are called a *Fourier transform* pair. These are the direct transform:

$$F(\omega) = \int_{-\infty}^{+\infty} f(x)\, \epsilon^{-j\omega x}\, dx; \tag{3-25}$$

and the inverse transform:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega)\, \epsilon^{j\omega x}\, d\omega \tag{3-26}$$

Equations (3-25) and (3-26) are the complex notation versions of the Fourier integral. Though equivalent noncomplex notation equations, corresponding to similar equations for the Fourier series, can be developed, we shall not do so as the complex notation is the easier to use.

Complex notation is helpful in Fourier transform useage because it leads to a pair of symmetrical equations. Thus, except for the factor $\frac{1}{2\pi}$ which comes from using the variable ($\omega = 2\pi f$) and the change of sign in the ($j\omega x$) exponent, the two equations are identical. This means that except for a change of scale, functions and their transform are interchangeable. The fact that the frequency-domain representation of a rectangular pulse (direct transform) is $(\sin x)/x$ indicates that to get a rectangular spectral distribution one has to start with a $(\sin x)/x$ time-domain function.

As an example in using Fourier transforms, let us consider a rectangular pulse of width t_0 and amplitude A, such as one of the pulses shown in Fig. 3-2A.

The time-domain function is:

$$f(t) = A,$$

$-\frac{t_0}{2} < t < +\frac{t_0}{2}$, and zero everywhere else.

The direct Fourier transform equation is:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)\, \epsilon^{-j\omega t}\, dt.$$

Substituting for $f(t)$ the value A between the limits of $\frac{t_0}{2}$ and $\frac{-t_0}{2}$, we have:

$$F(\omega) = A \int_{-\frac{t_0}{2}}^{+\frac{t_0}{2}} \epsilon^{-j\omega t}\, dt$$

$$= A\, \frac{(-1)}{j\omega}\, \epsilon^{-j\omega t} \Bigg]_{-\frac{t_0}{2}}^{+\frac{t_0}{2}}$$

$$= A \frac{1}{j\omega} \left(\epsilon^{j\omega \frac{t_0}{2}} - \epsilon^{-j\omega \frac{t_0}{2}} \right)$$

Substituting Euler's identity as discussed in Chapter 2, the result is:

$$F(\omega) = A t_0 \frac{\sin \pi f t_0}{\pi f t_0}, \tag{3-27}$$

where

$$\omega = 2\pi f.$$

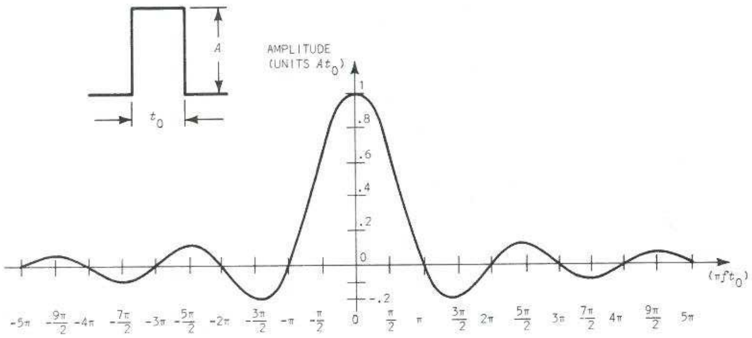


Fig. 3-6. Fourier transform of rectangular pulse.

Except for the multiplication factor of 2, expression (3-27) is identical to (3-22) which was developed through physical reasoning. As previously indicated, the factor of 2 was introduced while eliminating the energy distribution at negative frequencies, which is difficult to handle in a physically meaningful way. Equation (3-27) is plotted in Fig. 3-6. Comparison of Figs. 3-5C and 3-6 shows that the major difference between theory and practice is that in theory alternate lobes are of opposite phase, whereas in practice no such distinction is made.

A table of some common transform pairs and properties of Fourier transforms appears at the end of this chapter. Of major significance is the center-frequency-shift property as a function of complex modulation. Basically, when a time function is multiplied by a carrier at frequency ω_0, the spectrum of that function is shifted by ω_0. Thus, for the rectangular pulse, whose spectral distribution is shown in Fig. 3-6, the main lobe centered at $\pi f t_0 = 0$ moves from DC to ω_0 when the pulse is produced by turning on and off a sinusoidal carrier at ω_0. The frequency f in the angle $\pi f t_0$ becomes the difference frequency between the carrier at f_0 and the frequency of interest. This eliminates

pulsed RF

negative frequencies so long as $f_0 - f > 0$. Therefore, for a pulsed RF wave the spectral distribution as shown on a spectrum analyzer is symmetrical about the main lobe as in Fig. 3-6 rather than unidirectional as in Fig. 3-5C. This frequency shift also eliminates the factor of 2 from equation (3-22), correlating practice and theory.

An interesting and useful characteristic of Fourier integral analysis is that the results are valid not only for single pulse or transient phenomena but for pulse trains as well. Note that the shape of the curve generated by connecting the end points of the harmonic amplitudes in Fig. 3-5A and B is the same $(\sin x)/x$ as that of Fig. 3-5C. Therefore, all that is needed to reconstruct the harmonic amplitudes for a pulse train is a knowledge of how frequently to sample the continuous curve obtained by taking the Fourier transform of the pulse.

APPENDIX

1) EVALUATING FOURIER COEFFICIENTS

The basic Fourier series given by equation (3-1) is:

$$f(x) = \frac{a_0}{2} + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x)$$
$$\cdots + (a_n \cos nx + b_n \sin nx) \tag{3-28}$$

To determine the three coefficients a_0, a_n, and b_n, it is necessary to evaluate the following three integrals:

$$\int_{-\pi}^{+\pi} f(x)\,dx,$$
$$\int_{-\pi}^{+\pi} f(x)\cos nx\,dx, \tag{3-29}$$
$$\int_{-\pi}^{+\pi} f(x)\sin nx\,dx.$$

The evaluation of the integrals in (3-29) is performed by substituting the series (3-28) for $f(x)$ and integrating term by term. At first this appears to be an impossible task since (3-28) is an infinite series. However, in each of the three integrals, all the terms except one yield zero. This stems from the following basic integral relationships for sinusoids:

$$\begin{aligned} &\int_{-\pi}^{+\pi} \cos nx \cos mx \, dx = 0, \qquad m \neq n; \\ &\int_{-\pi}^{+\pi} \sin nx \sin mx \, dx = 0, \qquad m \neq n; \\ &\int_{-\pi}^{+\pi} \sin nx \cos mx \, dx = 0. \end{aligned} \tag{3-30}$$

The relationships in (3-30) are simply an expression of the fact that sines and cosines form an orthogonal set of functions.

Two special cases of the above are:

$$\begin{aligned} &\int_{-\pi}^{+\pi} \cos mx \, dx = 0 \\ &\int_{-\pi}^{+\pi} \sin mx \, dx = 0 \end{aligned} \tag{3-31}$$

Using (3-30) and (3-31), the integrals in (3-29) are easily evaluated as follows:

$$\begin{aligned} \int_{-\pi}^{+\pi} f(x) \, dx = &\int_{-\pi}^{+\pi} \frac{a_0}{2} \, dx + \int_{-\pi}^{+\pi} a_1 \cos x \, dx \\ &+ \int_{-\pi}^{+\pi} b_1 \sin x \, dx + \\ &\cdots \int_{-\pi}^{+\pi} a_n \cos nx \, dx \\ &+ \int_{-\pi}^{+\pi} b_n \sin nx \, dx. \end{aligned} \tag{3-32}$$

But from (3-30) and (3-31) all the terms except the first are zero, so:

$$\int_{-\pi}^{+\pi} f(x) \, dx = \int_{-\pi}^{+\pi} \frac{a_0}{2} \, dx = \pi a_0. \tag{3-33}$$

which is the same as equation (3-3) given as

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(x) \, dx.$$

To determine a_n we evaluate

$$\begin{aligned} \int_{-\pi}^{+\pi} f(x) \cos nx \, dx = &\int_{-\pi}^{+\pi} \frac{a_0}{2} \cos nx \, dx + \\ &\int_{-\pi}^{+\pi} a_1 \cos x \cos nx \, dx + \\ &\int_{-\pi}^{+\pi} b_1 \sin x \cos nx \, dx + \\ &\cdots \int_{-\pi}^{+\pi} a_n \cos nx \cos nx \, dx + \\ &\int_{-\pi}^{+\pi} b_n \sin nx \cos nx \, dx \end{aligned} \tag{3-34}$$

Based on (3-30) and (3-31), all the terms in (3-34) are zero except one, which leaves:

$$\int_{-\pi}^{+\pi} f(x) \cos nx \, dx = \int_{-\pi}^{+\pi} a_n \cos nx \cos nx \, dx,$$

which when evaluated leads to the result

$$\int_{-\pi}^{+\pi} f(x) \cos nx \, dx = \int_{-\pi}^{+\pi} a_n \cos^2 nx \, dx = \pi a_n. \tag{3-35}$$

Equation (3-35) is identical with that of equation (3-3):

$$a_n = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(x) \cos nx \, dx.$$

Similar reasoning leads to the result that:

$$b_n = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(x) \sin nx \, dx \qquad (3\text{-}36)$$

Equations (3-33), (3-35) and (3-36) are used in evaluating the coefficients in the Fourier series expansion.

2) THE IMPULSE FUNCTION

The time-domain unit impulse, designated delta of t, $\delta(t)$, is a function having infinitely narrow pulse width and unity area, so that its amplitude approaches infinity. Mathematically a unit impulse at time $t = 0$ has the property that:

$$\begin{aligned} \delta(t) &= 0, \; t \neq 0 \\ \delta(t) &\rightarrow \infty, \; t = 0 \end{aligned} \qquad \int_{-\epsilon}^{+\epsilon} \delta(t) \, dt = 1 \qquad (3\text{-}37)$$

where epsilon is an arbitrarily small time interval.

The area of the impulse is called its strength, so that a unit impulse has a strength of one. Naturally, it is not mandatory that all impulses have unit strength, any strength at all is possible. The unit impulse is, however, a convenient quantity to manipulate, so that other impulses are defined in terms of the unit impulse.

The impulse is not a practically realizable function, because real circuits cannot generate infinitely narrow arbitrarily large pulses. From a more fundamental point of view, the impulse is not realizable because the energy content of an impulse is infinite. This is because energy is proportional to the square of the impulse function. Thus the area of an impulse is

$$\int_{-\infty}^{+\infty} \delta(t) \, dt$$

and is finite, but the energy is given by

$$\int_{-\infty}^{+\infty} [\delta(t)]^2 \, dt,$$

which is infinite. This matter of infinite energy becomes clearer when the impulse is considered in the frequency domain. Thus, taking the Fourier transform of the unit impulse

$$F(\omega) = \int_{-\infty}^{+\infty} \delta(t) \, \epsilon^{-j\omega t} \, dt = 1.$$

The fact that the spectral distribution of an impulse is a constant means that the impulse has a constant energy per unit bandwidth at all frequencies. Hence, as we go to higher and higher frequencies, the energy increases without bound. What happens in real life is that the spectral density starts falling off at some arbitrarily high frequency, so that the total energy remains finite.

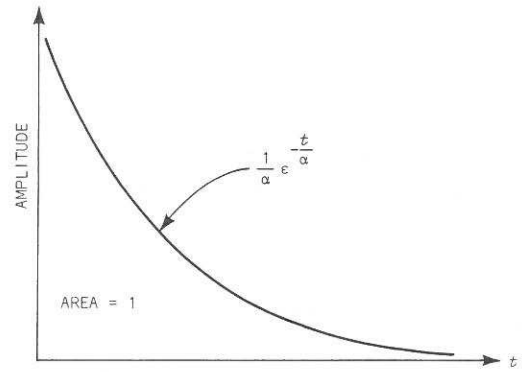


Fig. 3-7. Exponential prototype of impulse.

Since mathematical impulses do not exist in nature, it is of interest to investigate various approximations. The impulse can be approximated by any pulse such that the area remains constant as the width is decreased. A rectangular pulse of width τ and height $1/\tau$ has a constant unity area for all τ, and is therefore an acceptable representation of the unit impulse. The exponential function $\frac{1}{\alpha}\epsilon^{-\frac{t}{\alpha}}$, shown graphically in Fig. 3-7, is frequently used to represent an impulse. The area under this curve, as α goes to zero, is:

$$\text{Area} = \int_0^\infty f(t)\,dt = \frac{1}{\alpha}\int_0^\infty \epsilon^{-\frac{t}{\alpha}}\,dt = -\epsilon^{-\frac{t}{\alpha}}\Big]_0^\infty$$

Determining $\lim\limits_{\alpha\to 0}\int_0^\infty f(t)\,dt$:

$$\lim_{\alpha\to 0} -\epsilon^{-\frac{t}{\alpha}}\Big]_0^\infty = \lim_{\substack{\alpha\to 0\\ t\to 0}} \epsilon^{-\frac{t}{\alpha}} - \lim_{\substack{\alpha\to 0\\ t\to\infty}} \epsilon^{-\frac{t}{\alpha}} = 1. \qquad (3\text{-}38)$$

The exponential curve is, therefore, a good beginning shape for the unit impulse. The shape of the starting pulse really doesn't matter so long as the area remains constant as the pulse width is reduced.

As previously indicated, the pulse width can never be reduced to zero. However, when the pulse width is reduced to the point where the spectral distribution is constant over the frequency range of the circuits used, we have, for all practical purposes, generated an impulse.

reciprocal spreading

The flat frequency distribution of the impulse is connected with a property of the Fourier transform pair which is sometimes called *reciprocal spreading*: When one member of the transform pair is made narrower, the other spreads out and vice versa. Thus, for the case of the rectangular pulse, the first null occurs when the angle $\pi f t_0 = \pi$, which happens at the frequency $f = 1/t_0$. As the pulse width (t_0) is made narrower, the frequency width occupied by the main lobe in Fig. 3-6 gets wider and wider until, as the pulse width goes to zero, the peak of the main lobe is spread out over all frequencies.

Conversely, as the pulse width (t_0) is made wider, all of the energy gets more and more concentrated at the center frequency of the main lobe until, as the pulse width approaches infinity, the complete frequency distribution gets concentrated in an infinitesimally thin frequency band. This is, of course, a frequency impulse as discussed in connection with the Fourier transform of an infinitely long sinusoid. The frequency impulse is quite similar in its properties to the time impulse. All of the mathematics are almost identical except for the change of variable from t to f. This should not be surprising in view of the symmetry of the Fourier transform pair. Like its time-domain counterpart, true frequency-domain impulses do not exist in nature, since an infinitely long sinusoid is obviously impossible to generate. However, so long as the sinusoid exists for a long time, compared to the time constants of the circuits involved, it can be treated as if it were of infinite duration.

NAME	TIME DOMAIN	FREQUENCY DOMAIN	COMMENTS
1) DIRECT TRANSFORM	$f(t)$	$F(\omega) = \int_{-\infty}^{+\infty} f(t)\ \varepsilon^{-j\omega t}\ dt$	Fourier transform pair in complex notation.
2) INVERSE TRANSFORM	$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega)\ \varepsilon^{j\omega t}\ d\omega$	$F(\omega)$	
3) SUPERPOSITION	$f_1(t) + f_2(t)$	$F_1(\omega) + F_2(\omega)$	The Fourier transform of the sum of several time-domain functions is equal to the sum of the individual transforms. In this respect superposition is equivalent to addition.
4) DUALITY OF TRANSFORM PAIR	$f(t)$	$F(\omega)$	The transform pair differ from each other only in a scale change of 2π and in the sign of an exponent. Thus, if a time-domain shape (e.g., rectangle) gives a certain frequency shape, e.g., $(\sin x)/x$, then except for a scale change and sign inversion, a time shape like the frequency shape [$(\sin x)/x$ in time] will result in a frequency shape like the time shape (rectangular pulse in frequency domain).
	$F(t)$	$2\pi f(-\omega)$	
	$\frac{1}{2\pi}F(-t)$	$f(\omega)$	
5) CHANGE OF SCALE (RECIPROCAL SPREADING)	$f(\alpha t)$	$\frac{1}{\alpha}F\left(\frac{\omega}{\alpha}\right)$	This is related to the phenomenon which is sometimes referred to as reciprocal spreading. As one member of a Fourier transform pair narrows, the other spreads out. Futhermore, as the function spreads out the amplitude is reduced. This is a necessary consequence of the requirement that the total energy be the same whether the function is considered in time or frequency. Reciprocal spreading seems to be a universal property that applies to many reciprocal parameters. As the precision with which one parameter can be known is increased, the precision with which the other can be known is decreased. The best known expression of this type is the Heisenberg uncertainty principle in quantum mechanics which states that the product of the measurement accuracy of two reciprocal parameters, such as position and momentum, cannot be less than Planck's constant. Mathematically: $\Delta p \Delta q \approx h$ Heisenberg principle. This is similar to: $\Delta f \Delta t \approx 1$ Fourier theory.
6) EQUAL ENERGY OR PARSEVAL'S THEOREM	$\int_{-\infty}^{+\infty} [f(t)]^2\ dt$	$\frac{1}{2\pi}\int_{-\infty}^{+\infty} \lvert F(\omega)\rvert^2\ d\omega$	The energy computed by taking the integral of the square of the time-domain function or the integral of the square of absolute value of the frequency-domain function is the same. In other words, the time- and frequency-domain representations are equivalent as far as energy is concerned.
7) TIME SHIFT	$f(t-\tau)$	$\varepsilon^{-j\omega\tau}\ F(\omega)$	A change in position by $-\tau$ in the time domain introduces a phase shift of $\phi = -\omega\tau$ in the frequency domain. Conversely, introducing a phase shift in the frequency domain is equivalent to a delay in the time domain.
8) FREQUENCY SHIFT	$\varepsilon^{j\omega_0 t}\ f(t)$	$F(\omega-\omega_0)$	This is the inverse or dual of the time shift. The time operation is that of multiplication by another frequency, which in engineering is ordinarily called modulation. Thus, a shift in the frequency of a carrier that is pulsed on and off introduces a corresponding shift in the spectrum of the pulse.
9) CONVOLUTION	$f_1(t)\ f_2(t)$	$\frac{1}{2\pi}\int_{-\infty}^{+\infty} F_1(p)\ F_2(\omega-p)\ dp$	The spectrum of the product of two time-domain functions is the convolution of the individual spectra, where convolution is defined by the integral. The meaning of convolution is easier to understand by considering the inverse operation where the product of two spectra is equivalent to convolution in time. Convolution occurs when a signal is developed by sliding two functions past each other. Any scanning function, such as a filter sliding relative to a signal in spectrum analysis, involves convolution. The convolution of a function with an impulse results in the same shape as the original function. Hence, in spectrum analysis, if the scanning filter is assumed to be sufficiently narrow, the result of sliding the signal past the filter is the spectral characteristic of the signal.
	$\int_{-\infty}^{+\infty} f_1(\tau)\ f_2(t-\tau)\ d\tau$	$F_1(\omega)\ F_2(\omega)$	
10) DIFFERENTIATION	$\frac{d}{dt}\ f(t)$	$j\omega F(\omega)$	If the spectrum of $f(t)$ is given by $F(\omega)$, then the spectrum of the derivative of $f(t)$ is just $j\omega$ times that of $f(t)$. In other words, differentiation in the time domain means multiplication by $j\omega$ in the frequency domain.
11) INTEGRATION	$\int_{-\infty}^{t} f(t)\ dt$	$\frac{1}{j\omega}\ F(\omega)$	Integration is essentially the inverse of differentiation. Integration in the time domain means multiplication by $\frac{1}{j\omega}$ in the frequency domain.

Table 3-1.

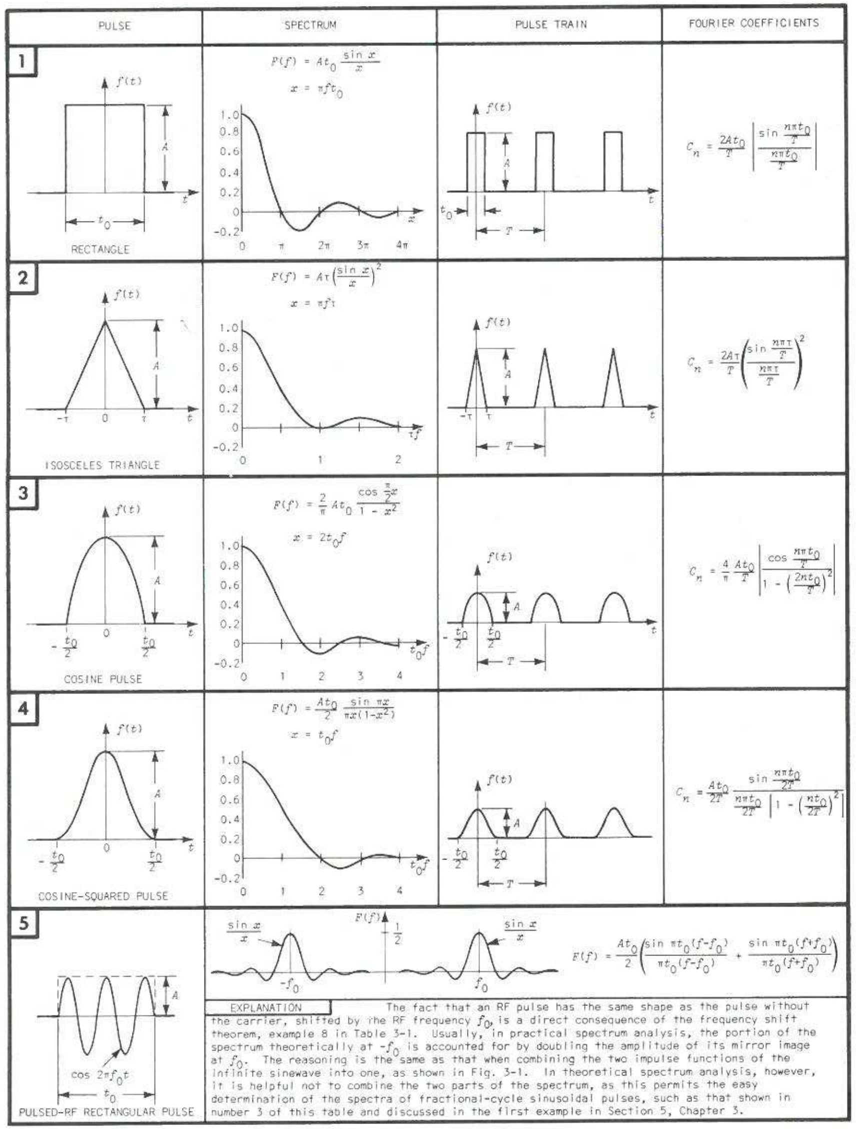


Table 3-2. Fourier transforms.

3) PROPERTIES OF FOURIER TRANSFORMS

A knowledge of some of the properties of Fourier transforms can be very helpful in spectrum analysis. Thus, for example, the fact that superposition holds for Fourier analysis can be helpful when interpreting complex spectra. Table 3-1 is a list of the more significant Fourier transform properties as found in spectrum analysis. No proofs are included. Some of these properties, such as convolution, have well-established names, while others may be found under different names in different tables.

4) TABLE OF FOURIER TRANSFORMS

Table 3-2 gives both graphical and mathematical relationships for time-domain to frequency-domain conversion.

5) EXAMPLES

A) USING ONE SPECTRUM TO DERIVE ANOTHER

Frequently, the spectral distribution of one waveform can be obtained by using the spectral distribution of another waveform. As an example, consider the half-cycle cosine pulse and the pulsed-RF rectangular pulse, numbers 3 and 5 respectively in Table 3-2.

The spectrum for a rectangular RF pulse of unity amplitude is from number 5 in Table 3-2:

$$F(f) = \frac{t_0}{2}\left(\frac{\sin\frac{1}{2}(\omega - \omega_0)t_0}{\frac{1}{2}(\omega - \omega_0)t_0} + \frac{\sin\frac{1}{2}(\omega + \omega_0)t_0}{\frac{1}{2}(\omega + \omega_0)t_0}\right) \qquad (3\text{-}39)$$

The product $\omega_0 t_0$ is the radian angle through which the carrier at ω_0 advances during the time t_0. For a half-cycle pulse, $\omega_0 t_0 = \pi$, since a full cycle is 2π radians. Substituting $\pi = \omega_0 t_0$ and cancelling terms results in:

$$F(f) = \frac{\sin \frac{1}{2}(\omega t_0 - \pi)}{\omega - \omega_0} + \frac{\sin \frac{1}{2}(\omega t_0 + \pi)}{\omega + \omega_0}. \tag{3-40}$$

The factor t_0 is eliminated by substituting $t_0 = \pi/\omega_0$, which leads to:

$$F(f) = \frac{1}{\omega - \omega_0} \sin\left(\frac{\pi\omega}{2\omega_0} - \frac{\pi}{2}\right) + \frac{1}{\omega + \omega_0} \sin\left(\frac{\pi\omega}{2\omega_0} + \frac{\pi}{2}\right). \tag{3-41}$$

Equation (3-41) can be further simplified by using the basic trigonometric identities:

$$\begin{aligned} \sin (A + B) &= \sin A \cos B + \cos A \sin B, \\ \sin (A - B) &= \sin A \cos B - \cos A \sin B. \end{aligned} \tag{3-42}$$

Using expression (3-42) in (3-41) we have:

$$\begin{aligned} F(f) = {} & \frac{1}{\omega - \omega_0}\left(\sin \frac{\pi\omega}{2\omega_0} \cos \frac{\pi}{2} - \cos \frac{\pi\omega}{2\omega_0} \sin \frac{\pi}{2}\right) \\ & + \frac{1}{\omega + \omega_0}\left(\sin \frac{\pi\omega}{2\omega_0} \cos \frac{\pi}{2} + \cos \frac{\pi\omega}{2\omega_0} \sin \frac{\pi}{2}\right) \end{aligned} \tag{3-43}$$

But, $\sin \frac{\pi}{2} = 1$ and $\cos \frac{\pi}{2} = 0$, hence (3-43) reduces to:

$$F(f) = -\frac{1}{\omega - \omega_0} \cos \frac{\pi\omega}{2\omega_0} + \frac{1}{\omega + \omega_0} \cos \frac{\pi\omega}{2\omega_0} \tag{3-44}$$

Combining terms we have:

$$F(f) = \left(\frac{1}{\omega + \omega_0} - \frac{1}{\omega - \omega_0}\right) \cos \frac{\pi\omega}{2\omega_0} \tag{3-45}$$

Combining the terms in the parenthesis by means of the common denominator, $\omega^2 - \omega_0{}^2$, results in:

$$F(f) = \frac{-2\omega_0}{\omega^2 - \omega_0{}^2} \cos \frac{\pi\omega}{2\omega_0} \tag{3-46}$$

Normalizing with respect to ω_0 by letting $\omega/\omega_0 = x$, and substituting π/t_0 for ω_0 we get:

$$F(f) = \frac{2t_0}{\pi} \frac{1}{1 - x^2} \cos \frac{\pi}{2} x \tag{3-47}$$

This is the same expression as that given in number 3, Table 3-2 with the amplitude A set equal to unity.

While it took a bit of algebra and trigonometry, this exercise demonstrates the fact that knowing a few key transforms it is possible to obtain others without going through a solution of the (sometimes difficult) integral equations.

B) THE SINE INTEGRAL, Si(x)

When dealing with continuous spectra, it is inappropriate to operate in terms of individual harmonics. The proper way to consider the spectral distribution is in terms of spectral density, which is a per-unit frequency difference, or bandwidth, quantity. Thus, in actual measurements, the wider the bandwidth of the measuring apparatus the more energy should be intercepted and the greater should be the output indication. But, most spectra do not have a flat frequency spectrum, so the output depends on where the apparatus passband intercepts the spectral distribution. For the rectangular pulse, for example, there would be considerable output when the measuring filter is tuned to the center of the main lobe and very little output when the filter frequency is at a null point. This variation in spectral density is of course determined by the area under any small portion of the spectral distribution curve, such as in Fig. 3-6.

To obtain an area it is necessary to integrate. This leads to the importance of the *sine integral* Si(x), where

$$\mathrm{Si}(x) = \int_0^x \frac{\sin x}{x}\, dx. \tag{3-48}$$

At $x = 0$, Si$(x) = 0$, since no area is intercepted at zero bandwidth. This is in agreement with previous reasoning which led to the conclusion that the amplitude of an individual spectral line, which has zero frequency width, is zero. As x increases toward the first zero crossing at π, Si(x) keeps increasing until, at $x = \pi$, Si$(x) = 1.85$. As x goes greater than π, Si(x) starts decreasing because the curve $(\sin x)/x$ is now negative. At $x = 2\pi$, Si(x) starts to increase again, oscillating back and forth every time x increases by π. In order to treat negative-going areas equally with positive-going areas, one needs either to take the integral of $\left(\frac{\sin x}{x}\right)^2$ or, as is more common, to determine the difference between two sine integrals. Thus, if one is dealing with an instrument bandwidth ΔF and one wishes to know the output around frequency f, one needs to determine the value of Si(x_2) - Si(x_1) where:

$$x_2 = \pi t_0 \left(f + \frac{\Delta f}{2}\right),$$

$$x_1 = \pi t_0 \left(f - \frac{\Delta f}{2}\right)$$

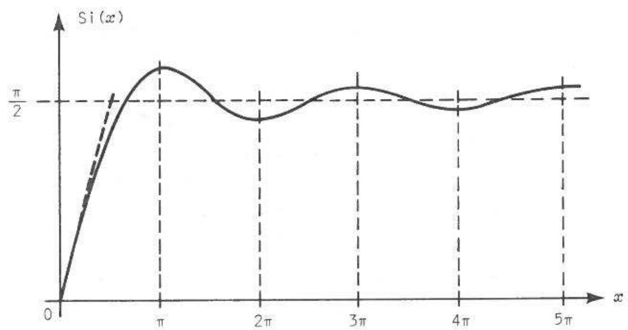


Fig. 3-8. A plot of the sine-integral of x.

The availability of sine integral tables can, therefore, be quite useful. Fortunately, this integral occurs in many communication problems so that tables are readily available. Fig. 3-8 is a plot of this integral.

C) RECTANGULAR PULSE ANALYSIS

Given a rectangular pulse, determine for its spectrum the position and amplitude of the first side lobe relative to the main lobe. The spectral distribution is given by the formula $F(f) = \frac{\sin x}{x}$, where $x = \pi f t_0$. A plot is given in Fig. 3-6. To determine the position of the various maxima, we differentiate and set equal to zero. Thus:

$$F(f) = \frac{\sin x}{x},$$

$$\frac{d}{dx} F(f) = \frac{x \cos x - \sin x}{x^2} = 0,$$

which means that maxima occur when

$$x = \tan x. \tag{3-49}$$

Observing the graph of $\frac{\sin x}{x}$ (Fig. 3-6), the maximum of the first side lobe occurs in the vicinity of $x = \frac{\pi}{2}$. By trial-and-error comparison of the value of tan x with x, the actual angle is found to be close to

$$x = 4.5 \text{ radians}, \tag{3-50}$$

which is 1.43π rather than the 1.5π estimated from Fig. 3-6. The relative amplitude is determined by substituting the appropriate value of x into the $(\sin x)/x$ equation. Thus, for the peak of the main lobe, $x = 0$. At small angles the sine of an angle is equal to the angle, so that

$$\lim_{x \to 0} \frac{\sin x}{x} = 1.$$

The amplitude of the first side lobe is $\frac{\sin 4.5}{4.5} = 0.2175$. The relative amplitude between the main lobe and first side lobe is

$$\frac{1}{0.2175} = 4.6, \text{ or } 20 \log 4.6 = 13.2 \text{ dB}. \qquad (3\text{-}51)$$

Very often this number is approximated as 13.4 or 13.5 dB. This comes from the approximation that the peak occurs at $x = 1.5\pi$ radians. For most applications this is well within the measurement accuracy. However, in precision measurements 13.2 dB should be used.

If the pulse width is $t_0 = 1\ \mu s$, what is the frequency width of the spectrum lobes?

The spectrum nulls occur at a spacing of $x = \pi$. Since $x = \pi f t_0$, this happens at frequency multiples of $f = 1/t_0$. Hence, the lobe width is $1/1\ \mu s = 1$ MHz.

What if the pulse width is increased to 10 μs? The lobe spacing then becomes $1/10\ \mu s = 100$ kHz. This is an example of reciprocal spreading, where, as the pulse width gets wider, the spectral width gets narrower.

What happens if the pulse consists of a 1-μs burst of a 1-GHz sinusoid? By the frequency shift theorem, number 8 in Table 3-1, the center frequency of the main lobe is shifted to 1 GHz. The spectral shape remains a $(\sin x)/x$ as before.

What are the characteristics of a 1-μs burst of a 500-kHz sinusoid? The first tendency is to say that the result is a $(\sin x)/x$ centered at 500 kHz. However, 1 μs is not sufficient time to establish the pulse shape of the gating oscillator when gating a 500-kHz carrier. As a matter of fact, we pass only a half cycle of 500 kHz during 1 μs. The result is the cosine pulse discussed previously in Example (A) and transform number 3 in Table 3-2.

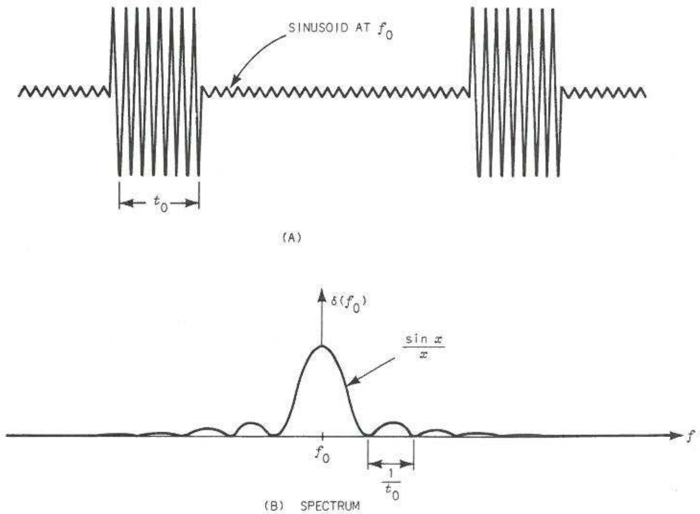


Fig. 3-9. Pulsed-RF with poor on/off ratio.

How is the spectrum of a burst of a 1-GHz sinusoid modified if the sinusoid does not turn off completely, as shown in Fig. 3-9?

If the 1-μs pulse width is a substantial part of the waveform cycle, we need to consider Fig. 3-9A as composed of two pulses, one short and large in amplitude and the other long and small in amplitude. If, as is usually the case, the burst occupies only a small part of the waveform period and is very much larger in amplitude than the remaining sinusoid, the sinusoid appears for all practical purposes as an uninterrupted CW signal. In either case, superposition applies, so that the waveform can be considered to be composed of two components. In the prevalent case of a large narrow pulse, the spectrum consists of the superposition of the $(\sin x)/x$ for the pulse and an impulse for the sinusoid, as shown in Fig. 3-9B.

4

MODULATION THEORY

In electronic communications, the message is usually not in a form suitable for transmission over the medium intervening between transmitter and receiver. The process whereby the original message is modified into an information-bearing transmittable signal is called *modulation.* Modulation theory is a vast subject that properly belongs as part of information theory. It is certainly not the intent, nor within the scope, of this volume to treat a subject of such complexity. Fortunately, the two forms of modulation (AM and FM), most often described in the frequency domain by measurement with spectrum analyzers, are also the most amenable to relatively simple mathematical analysis. Such important topics as pulse-code modulation (PCM) or time or frequency multiplexing will not be considered. Within the information-theory meaning of the word "modulation," we shall only consider amplitude modulation (AM) and angle modulation in the form of frequency modulation (FM).

To discuss modulation, it is necessary to state three definitions:

1) Carrier: the wave to which modulation is applied.
2) Modulating wave: the signal which contains the original message and is used to control some parameter of the carrier.
3) Modulated wave (modulated carrier): the final result of the modulation process after the modulating wave has affected the carrier. This is the wave, or signal, that is sent by the transmitter to the receiver.

None of these three waves, or signals, need be sinusoidal. For example, in pulse modulation, the carrier consists of a train of pulses some parameter of which, such as pulse height or pulse position, is controlled by the modulating wave. The modulating wave, generated by speech, is certainly far from sinusoidal. Finally the modulated wave, consisting of some complex combination of the two above, can be quite complicated. Nevertheless, the analysis that follows is based on sinusoidal waveforms. This is justified on the basis that in AM and FM the carrier is a sinusoid and that any modulating wave can be broken into an equivalent series of sinusoids by means of Fourier analysis. The ultimate reason for doing things this way is, of course, the tremendous simplification in the analysis.

A sinusoid, $A \sin \theta$, has two basic parameters that can be varied: the amplitude A and the angle θ. Let us begin by analyzing the effect of a changing amplitude.

amplitude modulation

The carrier in amplitude modulation (AM) is usually a sinusoid of the form

$$A \sin (2\pi Ft + \alpha), \tag{4-1}$$

where A is the carrier *amplitude*, F is the carrier *frequency* and α is the initial *phase* or just phase. We will assume that the information desired to be transmitted is also sinusoidal in nature and represented mathematically by the modulating wave:

$$B \cos 2\pi ft \tag{4-2}$$

What is meant by AM is that the amplitude of the carrier is made to vary in proportion to the modulating wave, generating a modulated wave of the form:

$$a = A\,(1 + m \cos 2\pi ft) \sin (2\pi Ft + \alpha), \tag{4-3}$$

where m is called the degree of modulation, or $100m$ is the *percentage modulation*, f is the *modulation frequency* and a is the *instantaneous amplitude.* Usually the word amplitude refers to a constant such as A in equation (4-1). A more correct name for a might be instantaneous value. However, a is an amplitude in the sense that its square is proportional to instantaneous power. Fig. 4-1 is a graphical representation of equation (4-3).

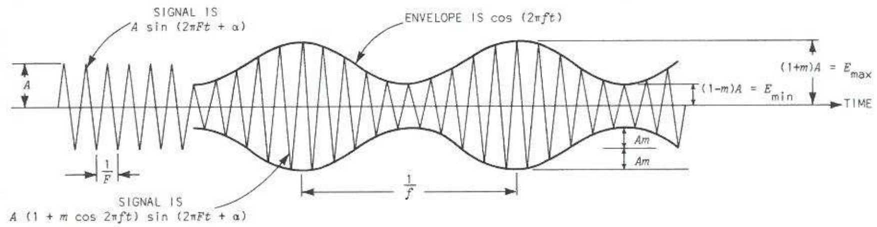


Fig. 4-1. Time-domain appearance of amplitude modulation.

In order to obtain the frequency-domain representation of an amplitude-modulated wave, it is necessary to disassociate the complex expression (4-3) into a sum of individual sinusoids. This is easily accomplished with the help of the trigonometric identity

$$\sin A \cos B = \frac{1}{2}\left[\sin (A + B) + \sin (A - B)\right]. \tag{4-4}$$

Letting $2\pi Ft + \alpha = A$ and $2\pi ft = B$ and substituting (4-4) into (4-3), we obtain

$$\begin{aligned} a &= A\left[1 + m \cos (2\pi Ft) \sin (2\pi ft + \alpha)\right] \\ &= \underbrace{A \sin (2\pi Ft + \alpha)}_{\text{carrier}} + \underbrace{\frac{Am}{2} \sin \left[2\pi(F+f)t + \alpha\right]}_{\text{upper sideband}} \\ &\quad + \underbrace{\frac{Am}{2} \sin \left[2\pi(F-f)t + \alpha\right]}_{\text{lower sideband}}. \end{aligned} \tag{4-5}$$

sidebands

From (4-5) it will be observed that an AM wave can be considered as consisting of the original carrier and two new components called *sidebands*. The sidebands are spaced on either side of the carrier with a frequency spacing equal to the modulating frequency f. The amplitude of the sidebands, relative to that of the carrier, is equal to half the percentage modulation, $m/2$. The frequency-domain representation of an AM wave is shown in Fig. 4-2.

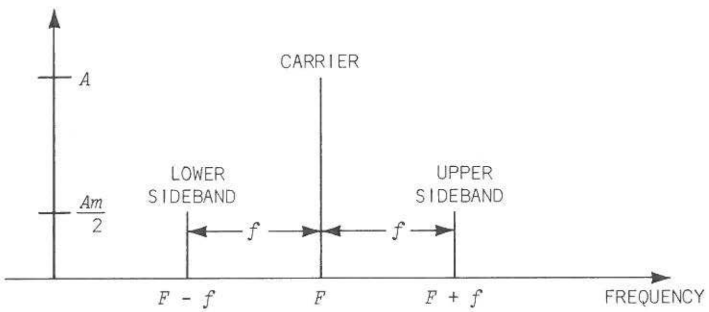


Fig. 4-2. Frequency-domain appearance of amplitude modulation.

energy

From equation (4-5) it is apparent that the carrier component of the AM spectrum is independent of the degree of modulation. Hence, an amplitude-modulated wave always contains more energy than the unmodulated carrier. At 100% modulation the sidebands are half as large as the carrier. This is the maximum relative amplitude that the sidebands can attain without overmodulation. Since energy is proportional to voltage amplitude squared, it follows that at 100% modulation each sideband contains one-quarter as much energy as the carrier. Thus, at maximum modulation the AM wave contains 50% more energy than the unmodulated carrier.

degree of modulation

What is of interest in determining amplitude modulation is the carrier frequency F, the modulating frequency f and the degree or percent modulation m. The frequencies are usually known beforehand, so the degree of modulation is the most frequent measurement. All three of the above parameters can theoretically be determined by use of either a time-domain oscilloscope or a frequency-domain spectrum analyzer.

In the time domain, illustrated in Fig. 4-1, the frequencies F and f are easily determined as the inverse of two simple time measurements, while the degree of modulation m is computed from a knowledge of the peak and null waveform amplitudes. Thus, the peak waveform amplitude is $E_{max} = (1+m)A$, while the null amplitude is $E_{min} = (1-m)A$. Calling the ratio of these some constant, K, we have

$$\frac{E_{max}}{E_{min}} = \frac{(1+m)A}{(1-m)A} = K, \qquad m = \frac{K-1}{K+1}. \tag{4-6}$$

In Fig. 4-1, K = 3, hence $m = \frac{3-1}{3+1} = 0.5$, or we have 50% of amplitude modulation.

While all the parameters can be obtained from time-domain measurements in theory, there are practical difficulties. Problems with time-domain measurements are:

1) Too high a carrier frequency for time-domain measurement. Sampling oscilloscopes have greatly reduced this problem.

2) Too small an amplitude level to be observed on an oscilloscope.

3) The complicated nature of the signal when more than one modulating frequency is involved.

4) Difficulty in determining the ratio of E_{max}/E_{min} at low percentages of modulation.

To alleviate problems such as those above, it is customary to make AM measurements by means of the frequency-domain spectrum analyzer. Here the carrier frequency F is obtained from the calibrated RF center frequency dial, the modulation frequency f is determined by measuring the frequency difference between the carrier and the sidebands, while the degree of modulation is determined by measuring the relative amplitude between carrier and sidebands and computing from

$$\% \text{ modulation} = m \cdot 100 = \frac{2 \cdot A_{\text{sideband}}}{A_{\text{carrier}}} 100. \qquad (4\text{-}7)$$

In the case of Fig. 4-2,

$$\% \text{ modulation} = 2 \cdot \frac{1}{4} \cdot 100 = 50\%.$$

Besides the standard carrier with double sideband AM, systems which eliminate the carrier, or the carrier and one sideband, are also utilized. These are:

1) Reduced- or suppressed-carrier AM.
2) Single-sideband (SSB) AM.
3) Vestigial-sideband AM.

The rationale for use of these systems stems from a desire to reduce transmitter power requirements and to utilize more efficiently the available frequency space.

suppressed-carrier AM

The suppressed-carrier signal, as the name implies, consists of two sidebands with a greatly reduced or attenuated carrier. Normally the carrier contains at least two-thirds of the transmitted power. The suppressed-carrier technique permits a reduction in transmitted power without reducing the size of the intelligence-bearing sidebands. The technique, while reducing transmitter power requirements, calls for a more complicated receiver design since the carrier has to be reinserted to avoid distortion.

single-sideband AM

In single-sideband transmission, the usual practice is to eliminate one sideband and the carrier, though elimination of only the sideband is also called single sideband. Eliminating one sideband cuts the transmitted spectral width in half, thus conserving frequency space. As in suppressed carrier, SSB requires a more complex receiver since the missing sideband has to be reinserted by generating a mirror image of the transmitted sideband.

vestigial-sideband AM

Sometimes, to ease network complexity, one sideband is merely reduced in amplitude rather than eliminated. This is particularly true when the information contains extremely low frequencies. Such an arrangement is called vestigial sideband. The best known example of vestigial-sideband transmission is in television. Here the vestigial sideband occupies about one-sixth the frequency space of the unattenuated sideband, thus conserving broadcast power and frequency space.

angle modulation

As in amplitude modulation, it is not necessary that the signals be sinusoidal for angle modulation. However, for ease of analysis we shall confine the analysis to sinusoids.

As the name implies, in angle modulation it is the angle rather than the amplitude of the sinusoidal carrier, $A \sin(2\pi Ft + \alpha)$, that is varied. There are essentially an infinite number of ways in which the angle of the carrier, $2\pi Ft + \alpha$, can be made to vary by the modulating wave, $B \cos 2\pi ft$. The two prevalent systems are phase modulation (PM) and frequency modulation (FM). In the former, the phase of the carrier is made to vary linearly with the modulating signal, while in the latter it is the frequency of the carrier which is made to vary in accordance with the modulating wave. For sinusoidal modulation, FM and PM are not much different since instantaneous frequency is the derivative of phase and the derivative of a sinusoid is a sinusoid, thus:

$$\text{Instantaneous frequency} = \frac{1}{2\pi}\frac{d\theta}{dt} \qquad (4\text{-}8)$$

When dealing with a single sinusoid of the form $A \sin 2\pi Ft$, the angle $\theta = 2\pi Ft$ and

$$\frac{1}{2\pi}\frac{d\theta}{dt} = \frac{1}{2\pi}\frac{d}{dt} 2\pi Ft = F \qquad (4\text{-}9)$$

This shows that the definition of instantaneous frequency, (4-8), is in agreement with the conventional understanding of the word.

phase modulation

Let us now consider PM and FM in more detail. For phase modulation, let the carrier be a sinusoid of the form $a = A \sin (2\pi Ft + \alpha)$ and let the modulating wave be represented by $B \cos 2\pi ft$. Phase modulation means that the phase of the carrier, which in unmodulated form is given by α, is modified by the modulating wave, resulting in a new phase of $(\alpha_0 + \Delta\alpha \cos 2\pi ft)$. The complete expression is

$$a = A \sin \left[2\pi Ft + (\alpha_0 + \Delta\alpha \cos 2\pi ft)\right] \qquad (4\text{-}10)$$

phase deviation

The quantity, $\Delta\alpha \cos 2\pi ft$, is called the *phase deviation* and is expressed in radians. To reiterate – in PM the phase of the carrier is made to vary in accordance with the instantaneous amplitude of the modulating waveform resulting in a modulated waveform as given by expression (4-10).

frequency modulation

Let us now consider FM, which differs little from PM as will be shown. Since FM is more commonly used, it will be examined in much greater detail.

Let the carrier be of the form $a = A \sin (2\pi Ft + \alpha)$ and the modulating waveform $B \cos 2\pi ft$. Frequency modulation means that the instantaneous frequency of the carrier is modified in accordance with the instantaneous amplitude of the modulating waveform.

Combining the definition of instantaneous frequency from (4-9) with the frequency of the carrier F and the form of the modulating waveform $\cos 2\pi ft$, we have for the frequency of the modulated waveform:

$$\frac{1}{2\pi}\frac{d\theta}{dt} = F + \Delta F \cos 2\pi ft. \qquad (4\text{-}11)$$

Equation (4-11) simply states that the instantaneous frequency of an FM signal is the sum of the carrier frequency and a term which has the form of the instantaneous amplitude of the modulating waveform, where θ stands for the phase of the modulated waveform which is given by $a = A \sin \theta$. To get the final equation for an FM signal, it is necessary to solve for θ. Integrating (4-11) we have

$$\int d\theta = \int 2\pi F\, dt + \int 2\pi\Delta F \cos (2\pi ft) dt,$$

or $\qquad (4\text{-}12)$

$$\theta = 2\pi Ft + \frac{\Delta F}{f} \sin 2\pi ft + \theta_0.$$

The final result is that the FM wave is of the form

$$a = A \sin \theta = A \sin \left(2\pi Ft + \frac{\Delta F}{f} \sin 2\pi ft + \theta_0\right). \qquad (4\text{-}13)$$

frequency deviation

The factor ΔF is called the *peak frequency deviation* while $\Delta F \cos 2\pi ft$ is the *frequency deviation.* Frequency deviation means deviation with respect to the carrier frequency F.

FM and PM differences

It will be observed that equations (4-13) for FM and (4-10) for PM are of the same form except for the factor $1/f$ in (4-13). The major difference between FM and PM, therefore, is that PM has greater deviations, at relatively high modulation frequencies, than FM. The difference between FM and PM is particularly noticeable for multitone modulation where the ratios of the deviations at different frequencies is different for FM and PM.

Fig. 4-3 shows the time-domain appearance of FM and PM. Note that the phase-modulated waveform has the appearance of frequency modulation of the integrated modulating waveform. This follows from the fact that instantaneous frequency is the differential of phase or, conversely, phase is the integral of instantaneous frequency. Since, for sinusoids, integration and differentiation merely involve a phase shift, it follows that, except for a change in deviation, FM and PM are the same for a sinusoidal modulating wave.

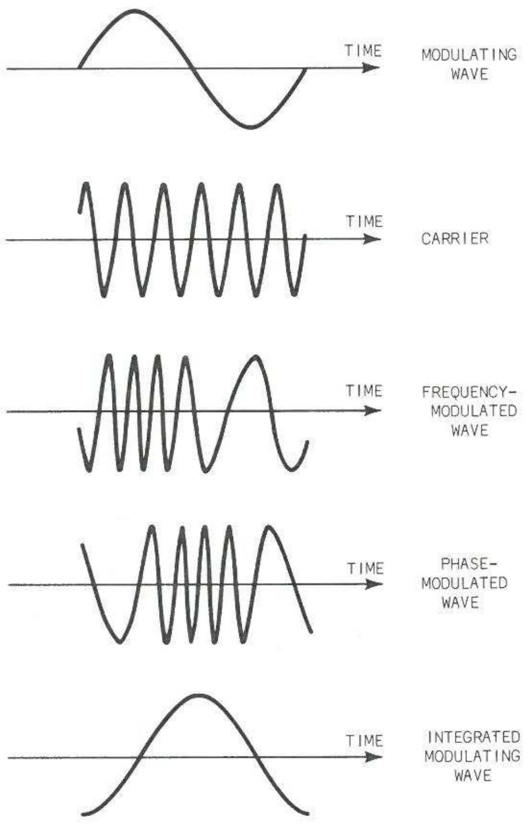


Fig. 4-3. Time-domain appearance of angle modulation.

Bessel functions

At this point, we digress for a discussion of Bessel functions which are necessary for the frequency-domain description of FM.

As discussed in Chapter 2, the circular trigonometric functions, $\sin\theta$ and $\cos\theta$ are the solution to the differential equation:

$$\frac{d^2y}{dt^2} + \omega^2 y = 0 \tag{4-14}$$

Similarly, Bessel functions are the solution of the differential equation:

$$\frac{d^2y}{dt^2} + \frac{1}{t}\frac{dy}{dt} + \left(1 - \frac{p^2}{t^2}\right)y = 0 \tag{4-15}$$

where p is a constant. While sinusoids, because of their long usage, appear simple and obvious, Bessel functions appear mysterious and forbidding. This need not be so. While we shall not derive the various Bessel-function relationships[1], just a discussion of the meaning of terminology can be tremendously helpful.

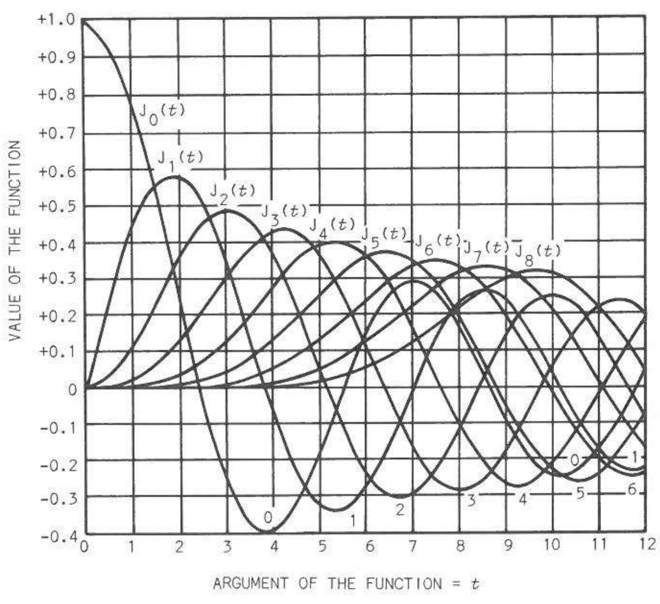
first kind

order

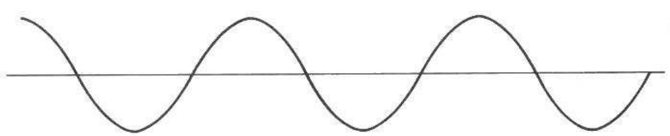
argument

Just as the solution of the circular equation (4-14) consists of two functions, a sine and a cosine, so the solution of Bessel equation (4-15) consists of two functions called Bessel functions of the first kind and Bessel functions of the second kind. We are mainly interested in Bessel functions of the first kind which are designated by the letter J. An equivalent statement for the circular trigonometric functions would be: We are only interested in cosines. There are two parameters, ω and t, associated with the cosine. Likewise, there are two parameters, called the *order* and *argument*, associated with Bessel functions. In Bessel-function language, $\cos\omega t$ would be called a circular trigonometric function of the first kind of order ω and argument t. $\sin\omega t$ would be a circular trigonometric function of the second kind of order ω and argument t. While in circular trigonometric functions the order and argument appear as a product, in Bessel functions they are separated as follows: $J_p(t)$ means Bessel function of the first kind of order p and argument t. Just as for the circular functions, there is no restriction on how large the order or argument of Bessel functions can get. These are also not restricted to integral values, though for FM applications the interest is in integer multiples of the order p.

[1] See, for example, Whittaker & Watson, *Modern Analysis*.



(A) BESSEL FUNCTIONS FOR THE FIRST 8 ORDERS



(B) cos ωt

Fig. 4-4.

Bessel functions are undulatory. But unlike the circular functions, the period, as measured between zero crossings, is not constant. Fig. 4-4A is a graph of the first eight orders of Bessel functions of the first kind. By contrast, cos ωt, Fig. 4-4B, has constant period and constant peak amplitudes.

The value of a Bessel function is much more difficult to calculate than that for a circular function. The result is usually obtained from an infinite series such as

$$J_p(t) = \frac{t^p}{2^p p!}\left(1 - \frac{t^2}{2(2p+2)} + \frac{t^4}{2 \cdot 4(2p+2)(2p+4)} \cdots \begin{array}{l}\text{for positive}\\ \text{integers of } p\end{array}\right) \qquad (4\text{-}16)$$

For the bessel function of the first kind, order zero, (4-16) becomes:

$$J_0(t) = 1 - \frac{t^2}{4} + \frac{t^4}{2 \cdot 4 \cdot 8} \cdots \qquad (4\text{-}17)$$

Fortunately, there are available many fine tables of Bessel functions[2], so that these can now be used almost as routinely as the circular trigonometric functions.

Bessel functions and the circular trigonometric functions are related. For example, at very large values of the argument t,

$$J_p(t) = \sqrt{\frac{2}{\pi t}} \cos\left(t - \frac{p\pi}{2} - \frac{\pi}{4}\right). \qquad (4\text{-}18)$$

That is, the larger the argument the closer does the Bessel function resemble a circular function. Bessel functions and the circular trigonometric functions are even more fundamentally related to each other, since it can be shown that:

$$\cos t = J_0(t) - 2\left[J_2(t) - J_4(t) + J_6(t) \cdots\right], \qquad (4\text{-}19)$$

$$\sin t = 2\left[J_1(t) - J_3(t) + J_5(t) \cdots\right]. \qquad (4\text{-}20)$$

The fact that a sinusoid can be expanded as a series of Bessel functions should cause no surprise. Bessel functions are orthogonal, hence, as discussed in Chapter 2, other functions including sinusoids are expandable as a series of Bessel functions.

[2] For example, see Jahnke and Emde, *Tables of Functions.*

Of major importance in FM theory is that a sinusoid with a sinusoidal modulation angle is expandable as a series of sinusoids with Bessel function coefficients. The formulas which are related to (4-19) and (4-20) are:

$$\cos(t \sin\theta) = J_0(t) + 2\left[J_2(t)\cos 2\theta + J_4(t)\cos 4\theta \cdots\right] \tag{4-21}$$

$$\sin(t \sin\theta) = 2\left[J_1(t)\sin\theta + J_3(t)\sin 3\theta \cdots\right] \tag{4-22}$$

$$\cos(t \cos\theta) = J_0(t) - 2\left[J_2(t)\cos 2\theta - J_4(t)\cos 4\theta + J_6(t)\cos 6\theta \cdots\right] \tag{4-23}$$

$$\sin(t \cos\theta) = 2\left[J_1(t)\cos\theta - J_3(t)\cos 3\theta + J_5(t)\cos 5\theta \cdots\right]. \tag{4-24}$$

Finally, an approximation useful in narrowband FM calculations is, for small arguments ($t < 0.5$), the zero order and first order Bessel functions of the first kind are related to the argument as follows:

$$\left.\begin{aligned} J_0(t) &\approx 1, \\ J_1(t) &\approx \frac{t}{2} \end{aligned}\right\} \quad \text{for } t < 0.5. \tag{4-25}$$

Let us now return to frequency modulation.

FM spectrum

The previously derived form of the FM wave is:

$$a = A \sin\left(2\pi Ft + \frac{\Delta F}{f}\sin 2\pi ft + \theta_0\right), \tag{4-26}$$

where: F is the carrier frequency,

f is the modulation frequency,

A is the carrier amplitude,

ΔF is the peak deviation, and

$\frac{\Delta F}{f}$ is called the *modulation index.*

Using the identities (4-21) through (4-24), it can be shown that the FM wave (4-26) is equivalent to an infinite series of sinusoids with Bessel coefficients as follows:

$$\begin{aligned} a &= A \sin\left(2\pi Ft + \frac{\Delta F}{f}\sin 2\pi ft + \theta_0\right) \\ &= A\left\{J_0\left(\frac{\Delta F}{f}\right)\sin\left(2\pi Ft + \theta_0\right) + J_1\left(\frac{\Delta F}{f}\right)\sin\left[2\pi(F+f)t + \theta_0\right]\right. \\ &\quad - J_1\left(\frac{\Delta F}{f}\right)\sin\left[2\pi(F-f)t + \theta_0\right] + J_2\left(\frac{\Delta F}{f}\right)\sin\left[2\pi(F+2f)t + \theta_0\right] \\ &\quad + J_2\left(\frac{\Delta F}{f}\right)\sin\left[2\pi(F-2f)t + \theta_0\right] + J_3\left(\frac{\Delta F}{f}\right)\sin\left[2\pi(F+3f)t + \theta_0\right] \\ &\quad \left. - J_3\left(\frac{\Delta F}{f}\right)\sin\left[2\pi(F-3f)t + \theta_0\right] + \cdots\right\} \end{aligned} \tag{4-27}$$

Equation (4-27) gives the frequency distribution, or spectrum, of a frequency-modulated wave. There are several important conclusions regarding the FM spectrum which can be drawn from (4-27):

1) The FM spectrum consists of a set of discrete sinusoids.

2) These sinusoids appear at carrier frequency F and sidebands on either side of the carrier spaced the modulating frequency f apart.

3) There is no end to the sindebands; theoretically, the FM spectrum has infinite frequency distribution.

4) The amplitudes of the carrier component and the various sidebands are determined by the product of the original carrier amplitude A and the value of a Bessel function. The order of the Bessel function corresponds to the sideband number counting the carrier as number zero. The argument of the Bessel functions is the modulation index $\frac{\Delta F}{f}$.

5) Since the amplitude of the carrier component is modified by the factor $J_0\left(\frac{\Delta F}{f}\right)$, it follows that the carrier component of the modulated wave is smaller in amplitude than the unmodulated carrier. As a matter of fact, the carrier component can actually go to zero. This is called a carrier null and happens when $J_0\left(\frac{\Delta F}{f}\right) = 0$. The first carrier null occurs at a modulation index of 2.4, as can be seen by the zero crossing of the $J_0(t)$ curve in Fig. 4-4A. These Bessel zeros are used in determining the frequency deviation as discussed in the section on measurements. FM is a constant-energy process where energy is removed from the carrier and supplied to the sidebands. Thus, the energy of an FM wave is constant regardless of the degree of modulation. This is in contrast to AM, where the carrier amplitude is constant and the modulation process adds energy to the wave.

Fig. 4-5 shows a typical FM spectrum. There are two important points that should be indicated. One is that while a spectrum analyzer, being insensitive to phase, will show the FM spectrum as in Fig. 4-5B, the actual spectrum is as shown in Fig. 4-5A. Here is shown the fact that the odd upper and lower sidebands are 180° out of phase with respect to each other. This is demonstrated in equation (4-27) by the alternating positive and negative signs associated with the odd numbered sidebands. This out-of-phase characteristic will be used in deriving the spectrum of combined AM and FM, and later in the section on applications as a means for differentiating between AM and narrowband FM. The second point is that, while theoretically the FM spectrum does go on *ad infinitum*, most of the energy is confined to a frequency band (plus and minus ΔF) around the carrier. This follows Bessel function theory where it can be shown that $J_p(t)$ diminishes rapidly when the order p is greater than the argument t. Since the order p is equal to harmonic number n, it follows that as the modulating frequency f gets smaller it takes more harmonics to cover the frequency range ΔF. Hence, as $f \rightarrow 0$ the frequency spectrum acquires sharp demarcation lines at plus and minus ΔF around F as shown in Fig. 4-5C.

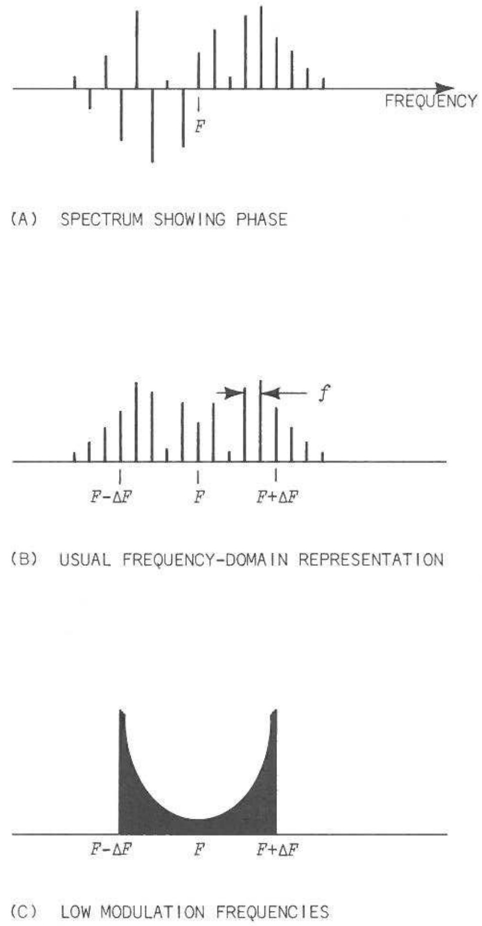


Fig. 4-5. FM spectrum.

multitone FM

In multitone AM, each modulating frequency can be treated individually as if the others were not there. Hence, the spectrum of multitone AM is just the sum of the individual single-tone spectra. In multitone FM, there is an interaction between the several modulating signal frequencies, creating additional sidebands than is apparent by treating each tone individually. The mathematics for multitone FM[3] can get quite complicated and will not be reproduced here. A major difference between the spectra of single-tone and multitone FM is that while in the former the sideband distribution is symmetrical about the carrier, in the latter it need not be. While an absolute rule is difficult to formulate, because of the complexity of the situation, it has generally been found that symmetrical modulating waveshapes create symmetrical spectra while unsymmetrical modulating waveshapes create unsymmetrical spectra. Thus, unless the modulating waveform is a pure sinusoid it is possible to get an unsymmetrical spectrum in FM. In multitone, as in singletone FM, the total energy is constant regardless of the degree of modulation. Hence, in multitone FM, as the number of sidebands is increased the carrier component is decreased.

combined AM and FM

incidental FM

Simultaneous AM and FM is usually an accidental, or incidental, phenomenon rather than a deliberate form of modulation. This form of modulation usually occurs when it is desired to obtain AM. Somehow the carrier oscillator frequency is pulled by the modulating signal, introducing a small amount of incidental FM along with the AM. The result is AM along with narrowband (low-modulation-index) FM at the same modulating frequency as the AM. Let us, therefore, consider the theoretical spectrum of AM combined with narrowband FM.

The AM spectrum consists of a carrier and two sidebands, as given by equation (4-5). The FM spectrum consists of a carrier and an infinity of sidebands, as given by equation (4-27). However, as previously indicated, the amplitude of the FM sidebands falls off very rapidly outside of the peak deviation interval $\pm\Delta F$. In narrowband FM, where ΔF is considerably less than the modulating frequency f, higher order sidebands fall off so rapidly that all but the first sideband can be ignored.

[3] See, for example, Giacoletto, "Generalized Theory of Multitone AM and FM," *Proc IRE*, July, 1947.

Thus, the spectrum is given by

$$a = A\left\{J_0\left(\frac{\Delta F}{f}\right)\sin\left(2\pi Ft + \theta_0\right) + J_1\left(\frac{\Delta F}{f}\right)\sin\left[2\pi(F+f)t + \theta_0\right] - J_1\left(\frac{\Delta F}{f}\right)\sin\left[2\pi(F-f)t + \theta_0\right]\right\} \tag{4-28}$$

which is simply equation (4-27), ignoring all but the first sideband. While narrowband FM differs from AM in that one sideband is 180° out of phase with respect to the other sideband, these are difficult to distinguish on a spectrum analyzer since the spectrum analyzer is insensitive to phase. A technique for distinguishing between narrowband FM and AM is discussed in the section on applications. Combined AM and narrowband FM, however, has a distinctive spectrum which is easily identified. Consider, for example, the case illustrated in Fig. 4-6. Here are shown the spectra of AM, narrowband FM and a combination of the two. While it is true that the spectrum analyzer is insensitive to phase, the displayed combined spectrum has to be considered

superposition

in accordance with the principle of superposition where phase has to be accounted for. The result, shown in Fig. 4-6C, is a distinctive spectrum where one sideband is larger than the other.

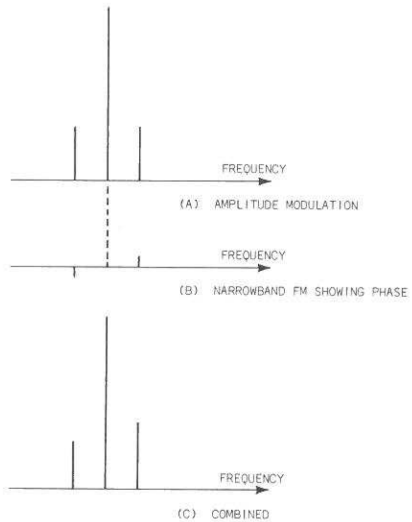


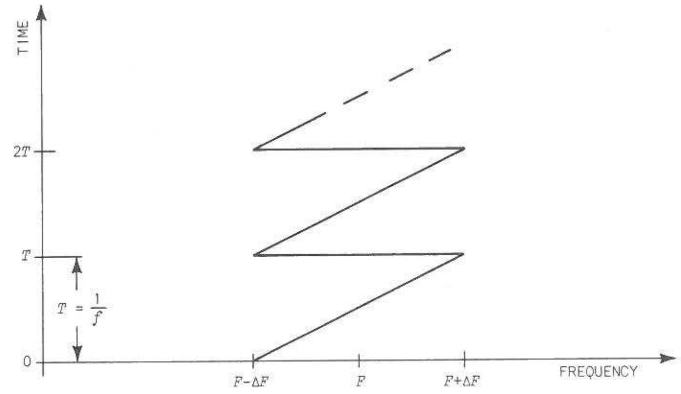
Fig. 4-6. Combining AM and narrowband FM.

One point that needs clarification is that while the sidebands are added the carrier is shown as constant amplitude. This is because the AM and FM spectra actually share the same carrier. In the FM spectrum, the carrier is shown dotted, because in narrowband FM there is so little energy in the sidebands that the carrier may be assumed unaffected, and of course in AM, the carrier is unaffected regardless of the percentage modulation. It is conceivable that the equipment is so constructed that the carrier level is affected by the incidental FM. This can only be determined from a knowledge of the actual equipment. In the absence of such knowledge we will assume that the carrier remains unaffected, which is the prevalent case.

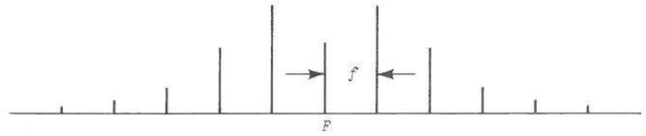
generating the FM spectrum

Many people have difficulty in accepting the validity of spectrum-analyzer derived results because many of these results seem contrary to common sense. While this problem was already considered in a general way in Chapter 2, it is worth while to resolve the specific case of FM. FM is of special interest because nowhere is the paradox between the spectral distribution and common sense more obvious than in this case.

Consider linear frequency modulation such as obtained from a sinusoidal carrier modulated with a sawtooth signal. The time-frequency relationship for the FM wave is shown in Fig. 4-7A. The carrier at a frequency F is made to vary its frequency in a linear fashion from $F - \Delta F$ to $F + \Delta F$, where ΔF is what we normally call the peak deviation. This process is repeated every T seconds or the FM rate is $f = 1/T$. It does not require any knowledge of FM theory to conclude that during the time interval T the FM signal goes through every frequency between $F - \Delta F$ and $F + \Delta F$. Furthermore, it is apparent from Fig. 4-7A that there are no frequency components outside the $F - \Delta F$ to $F + \Delta F$ interval. This rationale is behind the highly popular sweep-testing technique. In sweep testing the transfer characteristics of circuits, such as filters, are obtained by feeding a constant-amplitude FM signal into the circuit under test and observing the detected output. Since it is assumed that all frequency components are equally present at the circuit input within the limits of FM deviation, it follows that any variation in output amplitude is due to the circuit. Hence, the circuit characteristic is easily obtained.



(A) TIME-FREQUENCY PLOT FOR LINEAR FM



(B) TYPICAL SPECTRUM OF FM SIGNAL

Fig. 4-7.

Consider now a standard frequency-domain analysis of the FM signal. This result is shown in Fig. 4-7B. Here, the theory indicates that the FM wave consists of specific frequency components and nothing in between. These sinusoids, that appear to compose the FM wave, consist of a carrier component and sidebands spaced on either side of the carrier with a frequency spacing of $f = 1/T$. The two descriptions of the FM wave given in Figs. 4-7A and 4-7B appear to be in conflict in two areas —

1) Whereas Fig. 4-7A indicates that all frequencies between $F - \Delta F$ and $F + \Delta F$ are present, Fig. 4-7B indicates that there is only energy at very specific frequencies and nothing everywhere else.

2) While it is clear from Fig. 4-7A that no energy exists outside the frequency range $F - \Delta F$ to $F + \Delta F$, it is equally obvious from standard FM theory that there is no end to the sidebands.

So we have the paradox that the spectrum analyzer seems to indicate that there is nothing where logic says there should be something and, conversely, the spectrum analyzer indicates energy at frequencies where logically there should be nothing. Actually, both interpretations are correct because they apply to different circumstances.

First it should be recognized that for a meaningful, practical discussion it is necessary to consider the behavior of circuits because, in the final analysis, there is only one way to determine whether there is or isn't energy at a specific frequency — this is by means of a measurement using real equipment made up of circuits. We shall, therefore, consider the question — What is the difference between the circuits implied in the two approaches shown in Figs. 4-7A and 4-7B, and why do these different circuits give different frequency-domain results for FM?

As discussed elsewhere, there is only one way in which a spectrum analyzer is made to resolve or display individually the separate frequency components of a signal — this is by making the resolution bandwidth narrower than the frequency separation between signals. Hence, the difference is that Fig. 4-7A implies a relatively wideband circuit while Fig. 4-7B implies a relatively narrowband circuit. These circuits give different results because in one case the transient response is negligible while in the other case much of the output is due solely to the transient.

As discussed in Chapter 2, "Response of Circuits to Signals," a transient response need not be of short time duration. When dealing with high-Q narrowband circuits, the transient can be quite long. Hence, it is logical that the transient response of a narrowband filter should contribute more to the total output than is the case for a wideband filter. Now, consider the fact that the stimulus to the filter is repeated once every T seconds. This means that the transient response, whatever its characteristics, is repeated at intervals of T seconds. If the filter has a time constant such that the transient response does not die down too much in T seconds, it follows that the transient output never disappears since it is regenerated in a shorter time interval than it takes to die out. Time constants of filters are basically proportional to the inverse of the bandwidth. Hence, if the filter bandwidth is narrow enough to separate the several frequency components, meaning that the bandwidth is less than the FM rate f, it follows that the time constant is on the order of T and the transient is reconstituted faster than it dies out.

We shall not go through the derivation[4] of the network response leading to the remarkable fact that the total output of a narrowband filter with an FM input has energy only at discrete frequencies. While the input consists of a time-variable signal going through all the frequencies between $F - \Delta F$ and $F + \Delta F$ the output of a narrowband filter, consisting of the combined transient and steady-state response and averaged over one FM cycle of interval T, contains no energy except at the frequencies indicated by accepted FM theory. At all frequencies, except the very special ones, currents flow in the filter in such a way that on the average there is no energy transfer. Furthermore, if the narrowband filter is outside the frequency range of the input it is still possible to get an output. This is because a filter will respond with a transient to an input outside its frequency range. Normally, the transient dies down very quickly and can be ignored. However, if the stimulus is repeated at a fast enough rate and in appropriate synchronism with the filter frequency, one gets what looks like a continuous input. This is analogous to a swing pushed at a rate in synchronism with its natural frequency, resulting in continuous large oscillations.

When the filter bandwidth is large, as compared to the FM rate, the transient response is negligible so that the output has the same frequency characteristics as the input.

As was amply discussed in Chapter 2, it is not our intention to resolve the question of whether the spectral components are a part of the signal or are generated by the circuit. The important thing to remember is that *real, physically realizable, linear, time-invariant circuits behave as if spectral components exist*, and this is what we wish the spectrum analyzer to show.

---

[4]See Harvey, et al., "The Component Theory of Calculating Radio Spectra with Special Reference to FM," *Proc IRE*, June, 1951.

EXAMPLES

1) An amplitude modulated wave has a spectrum consisting of a carrier and two sidebands which are one one-hundredth the size of the carrier – What is the percentage modulation? From equation (4-7),

$$\%\ \text{modulation} = \frac{2 \cdot A_{\text{sideband}}}{A_{\text{carrier}}} 100 = \frac{2 \cdot \frac{1}{100}}{1} 100 = 2\%$$

2) Would this be a routine measurement on an oscilloscope? From equation (4-6), $m = \frac{\text{K-1}}{\text{K+1}} = \frac{2}{100}$, resulting in

K = 1.04. An amplitude ratio of 1.04 is very difficult to measure on an oscilloscope.

3) Given a wave which is frequency modulated at a 10-kHz rate. The spectrum shows the first carrier null. What is the deviation? The first carrier null occurs at a modulation index of 2.4; $\frac{\Delta F}{f} = 2.4$, $\Delta F = 2.4 \cdot 10 = 24$ kHz.

4) An FM spectrum shows a 10-kHz sideband spacing and the following relative amplitudes of its components: Carrier, one; first sideband, zero; second sideband, one; third sideband, one; fourth sideband, six tenths. What is the deviation? From Fig. 4-4A, $J_1(t)$ has zeros at an argument of about 3.8, 7, 10.2 $\cdots$. At only one of these does $J_0(t)$, $J_2(t)$ and $J_3(t)$ have the same magnitude – at a modulation index of 3.8. Here the various magnitudes are:

$$J_0(3.8) \cong -0.4,$$

$$J_1(3.8) \cong 0,$$

$$J_2(3.8) \cong 0.4,$$

$$J_3(3.8) \cong 0.4,$$

$$J_4(3.8) \cong .25;$$

these are in the ratios of $\frac{1}{1}, \frac{0}{1}, \frac{1}{1}, \frac{1}{1}, \frac{.625}{1}$.

Hence, the modulation index is 3.8 and the deviation is $3.8 \cdot 10 = 38$ kHz.

5) Given narrowband FM at a 10-kHz rate, the sidebands are one-fiftieth the amplitude of the carrier. What is the deviation? From equation (4-25),

$$J_0(t) \cong 1,$$

$$J_1(t) \cong \frac{t}{2} = \frac{1}{50}, \text{ and } t = \frac{1}{25}$$

The deviation is

$$\frac{1}{25}\ 10\ \text{kHz} = 400\ \text{Hz}.$$

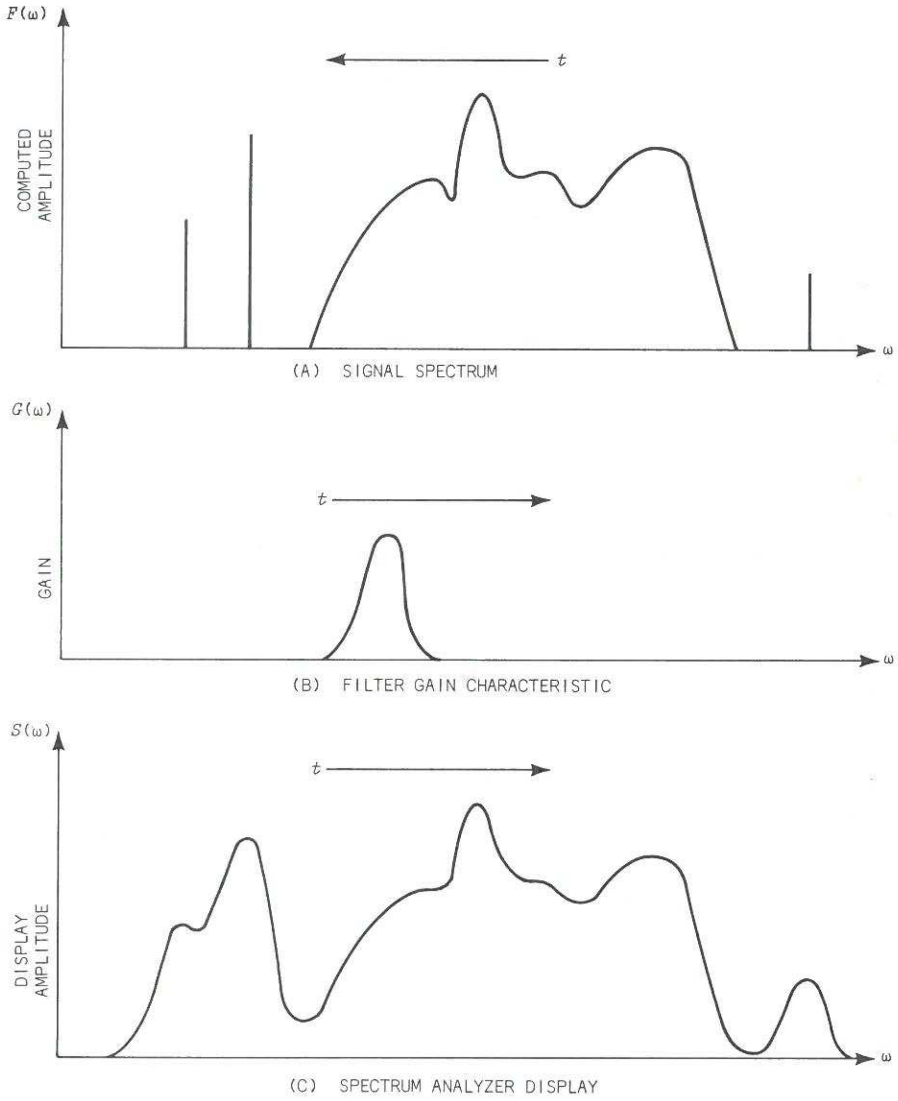


Fig. 5-1. Sweeping-signal spectrum-analyzer spectrum representation.

5

THE SWEEPING-SIGNAL SPECTRUM ANALYZER

As discussed in Chapter 1, the superheterodyne, or sweeping-signal spectrum analyzer, operates on the principle of the relative movement in frequency between the signal and a filter. The important parameter is the relative frequency movement. It does not matter whether the signal is stationary and the filter changes frequency nor whether the filter is stationary and the signal is made to change frequency.

frequency translation

Fig. 5-1 shows the spectral representation obtained in such a system. Fig. 5-1A represents a spectrum composed of three discrete-frequency CW signals and a continuous dense spectrum in the middle. This spectrum is passed through a filter having the gain characteristic shown in Fig. 5-1B. The filter and spectrum have a relative frequency translation as indicated by the arrows of opposite sense. The resultant display, shown in Fig. 5-1C, has the units of frequency, ω, but takes a real time, t, to occur. Some of the fine detail of the theoretical spectrum, shown in 5-1A as a Fourier transform $F(\omega)$, is lost in 5-1C because of the finite frequency width and, hence, resolution of the filter. As the resolution filter gets narrower, the ideal and actual spectral representations get more alike until, when the filter has zero bandwidth (in effect, an impulse function), the ideal and actual representations become the same. The transformation of the ideal spectrum into the actual spectral representation by the relative frequency translation between filter and signal is known as *convolution*. Convolution was previously discussed in Chapter 3, and a further discussion will be found in the appendix to this chapter.

convolution

While the basic operation of the system is apparent from Fig. 5-1, there are many ramifications, particularly with regard to the speed of relative-frequency translation, which are not at all obvious. Let us now consider some of the details of the sweeping-signal spectrum-analyzer system.

CW RESPONSE

While there are many possible configurations (e.g., swept IF or swept front end) as discussed in another chapter, all spectrum analyzers of the type under discussion contain a mixer, sweeping oscillator, and resolution filter. The simplest arrangement, which is sufficient for the purpose of theoretical discussion, is shown in Fig. 5-2. The time/frequency diagram for this system is shown in Fig. 5-3. Here it was assumed that system operation is based on a mixer output composed of the difference frequency between local oscillator and signal. Likewise, it was assumed that the signal is composed of two discrete frequency components. The signal components at frequencies f_1 and f_2 are shown as straight lines having infinitesimal frequency width and infinite time duration.

A constant-frequency signal is converted to a frequency sawtooth by combining it in a mixer with a frequency sawtooth from the swept local oscillator. In our example, it was assumed that the mixer output consists of the difference frequency between the local-oscillator frequency sawtooth and the input. Other combinations, such as the sum of the frequencies, lead to similar diagrams. The display consists of pulses whose time position is determined by the time of intersection of the filter passband and the sweeping signal, and whose width is equal to the time interval during which the sweeping-signal frequency is within the filter passband. The bursts or pulses generated by the relative translation of signal and filter are pseudo impulses representing the frequency-domain characteristics of the signal. While the *time position* of these pulses represents the input signal frequency and is determined by the incoming signal, the *width* τ of these pulses is determined solely by the spectrum-analyzer parameters. The width τ is equal to the time that the sweeping-signal frequency is within the passband of the filter, and from simple geometrical considerations is:

pulse width

$$\tau = \frac{B}{D} T \qquad (5\text{-}1)$$

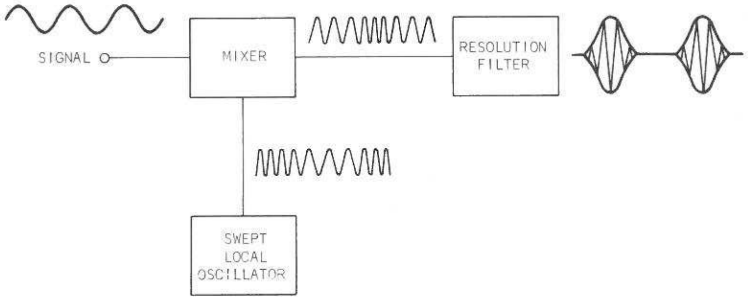


Fig. 5-2. Basic spectrum-analyzer block diagram.

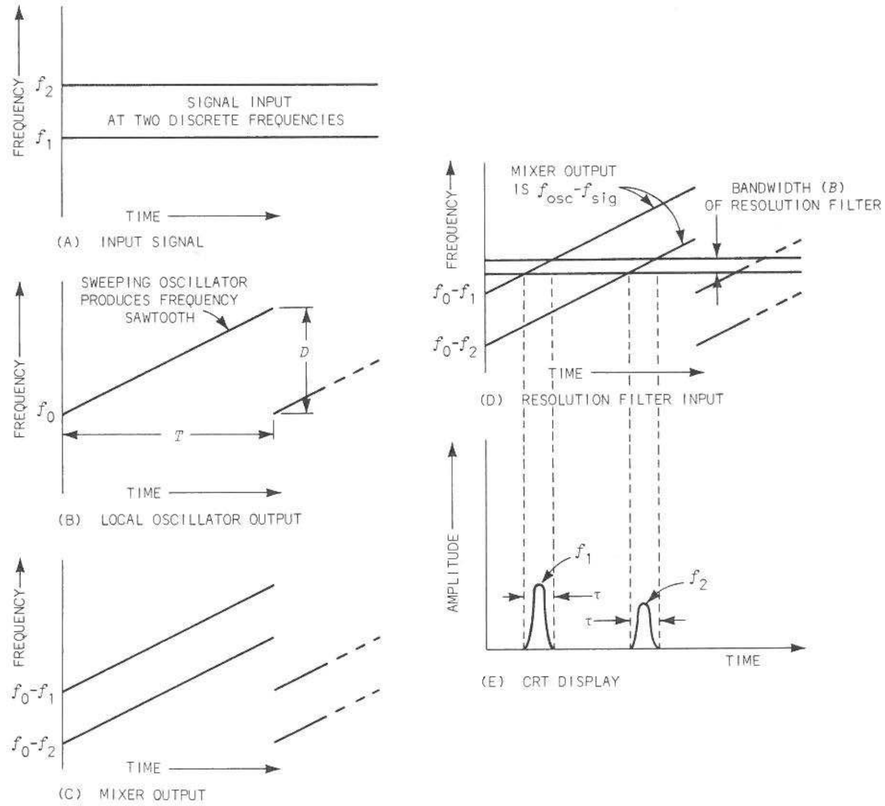


Fig. 5-3. Time/frequency diagram, sweeping-signal spectrum analyzer.

pulse-width parameters

The sweep time T is the time it takes the electron beam to traverse the horizontal width of the CRT. Hence the physical width of τ in inches or centimeters does not change with changing T. The actual time duration of τ, however, given by equation (5-1), is directly determined by the sweep time T. At low sweep time T, or with narrow resolution-filter bandwidth B, or with large dispersion D, the pulse width τ can become quite small. For example, a full-screen sweep time of 1 ms (100 μs/div), a resolution bandwidth of 10 kHz and a dispersion of 10 MHz result in a burst at the filter output which is only 1 μs wide. Such a narrow pulse cannot be passed by a 10-kHz-wide filter without distortion. As with any pulse that is passed through a filter of insufficient bandwidth, the output is of lesser amplitude and greater time duration than the input, as illustrated in Fig. 5-4. Since the distorted response is what appears on the CRT screen, the loss in amplitude shows a loss in sensitivity, and the apparent widening of the resolution bandwidth shows a loss in resolution. Analytical expressions, relating the amount of loss to the sweep time T, to actual resolution bandwidth B and to dispersion D, have been developed by many people. Some of the results are based on convolution techniques using the $\epsilon^{-\frac{1}{4}}$ bandwidth as the standard[1], while others calculate the transient response using a 3-dB standard bandwidth along the lines discussed in Section 4.6, Volume XI of the Radiation Laboratory Series[2]. When the differences in bandwidth are accounted for, the final results are essentially the same for both methods. The ratio of apparent resolution bandwidth R to actual bandwidth B is:

$$\frac{R}{B} = \left[1 + 0.195\left(\frac{D}{TB^2}\right)^2\right]^{\frac{1}{2}}, \qquad (5\text{-}2)$$

where B is the 3-dB bandwidth.

[1] Batten, et al., "The Response of a Panoramic Receiver to CW and Pulsed Signals," *Proc. IRE*, June, 1954.

Chang, "On the Filter Problem of the Power Spectrum Analyzer," *Proc. IRE*, August, 1954.

[2] *Spectrum Analyzer Techniques Handbook*, Polarad Electronics Corp.

"Spectrum Analysis," *Application Note 63*, Hewlett Packard.

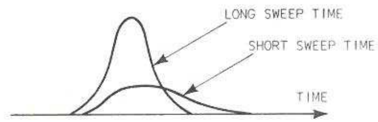


Fig. 5-4. Resolution distortion for short sweep time.

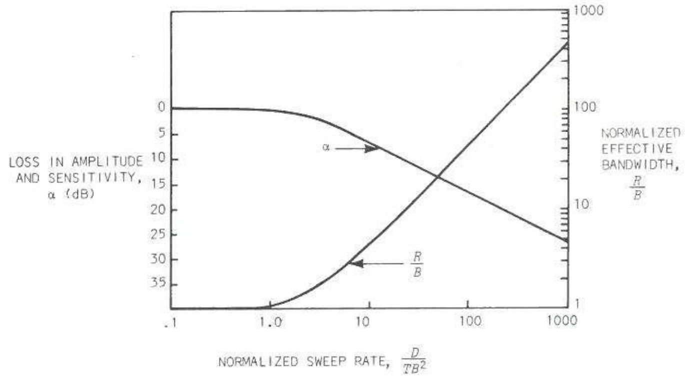


Fig. 5-5. Loss in sensitivity and resolution as a function of sweep rate.

The amplitude loss factor, α, is

$$\alpha = \left[1 + 0.195\left(\frac{D}{TB^2}\right)^2\right]^{-\frac{1}{4}}, \qquad (5\text{-}3)$$

where B is the 3-dB bandwidth.

phase response

Fig. 5-5 is a plot of equations (5-2) and (5-3). While these equations are theoretically correct, they do not necessarily correspond to the behavior of actual equipment. The discrepancy between theory and practice arises because the theoretical results are based on a filter having a gaussian amplitude response and linear phase response which in practice is not necessarily the case. Differences between theoretical and actual phase response can be particularly important. As a matter of fact, with appropriate phase response one can achieve pulse compression, such as utilized in chirp radar, so that resolution need not be degraded at high sweep rates[3].

[3] W. R. Kicheloe, Jr., "The Measurement of Frequency with Scanning Spectrum Analyzers," Report SEL-62-098, Stanford Electronics Laboratories.

While pulse-compression spectrum analyzers are presently economically unfeasible, the theory is nevertheless sound. Thus, while equations (5-2) and (5-3) are useful as a general guide to spectrum-analyzer performance, there is no substitute for experimental data on the actual equipment in question.

option of variables

In operating a spectrum analyzer, it is usually the practice to increase the sweep time T until the display is no longer distorted. There are, however, situations when this is not possible. Such a situation might arise when the signal under investigation has a limited time duration so the complete analysis must be performed in less than a specified time interval. Under such circumstances, sweep time T and dispersion D are usually fixed by the signal parameters and the only variable is the resolution bandwidth B.

Which value of actual bandwidth B results in minimum displayed resolution R? This is easily obtained[4] by differentiating (5-2) and letting $\frac{dR}{dB}$ equal zero. Thus:

$$\frac{dR}{dB} = B - \left[\frac{0.195}{B^3}\left(\frac{D}{T}\right)^2\right] = 0. \tag{5-4}$$

From which it follows that:

optimum resolution

$$B_o = \sqrt{\frac{1}{2.27}\frac{D}{T}}\ , \tag{5-5}$$

and

optimum resolution bandwidth

where B_o is the optimum bandwidth which, at a given setting of T and D, results in minimum (or optimum) displayed resolution bandwidth R_o.

When B_o is substituted back into equation (5-2), the result is that $R_o = \sqrt{2} \cdot B_o$. Theoretically, then,

$$R_o \cong \sqrt{\frac{D}{T}}\ ; \tag{5-6}$$

[4] Engelson & Long, "Optimizing Spectrum Analyzer Resolution," *Microwaves*, December, 1965.

any other value of B results in a larger R. R_o and B_o are generally known as optimum resolution and optimum resolution bandwidth respectively. While the optimum resolution is generally proportional to the square root of the ratio of dispersion to sweep time, the proportionality constant need not necessarily be unity as given by (5-6). Equation (5-6) is based on a gaussian, linear phase response filter which is not usually the case in practice.

PULSED SIGNALS

$\frac{\sin x}{x}$ energy distribution in frequency spectrum

In Chapter 3 it was shown that the theoretical spectrum of a train of pulses has a $\frac{\sin x}{x}$ shape as shown in Fig. 5-6. This $\frac{\sin x}{x}$ curve can be interpreted in two ways. One interpretation is that the $\frac{\sin x}{x}$ is the envelope formed by the locus of the end points of the fundamental and harmonic sinusoids which combine to produce the pulse train. The display consists of a set of vertical lines each representing a sinusoid. If the spectrum-analyzer dispersion were so adjusted as to show only one of these lines on the CRT, the observer would have no way of knowing that the input to the analyzer is not a single sinusoidal CW signal.

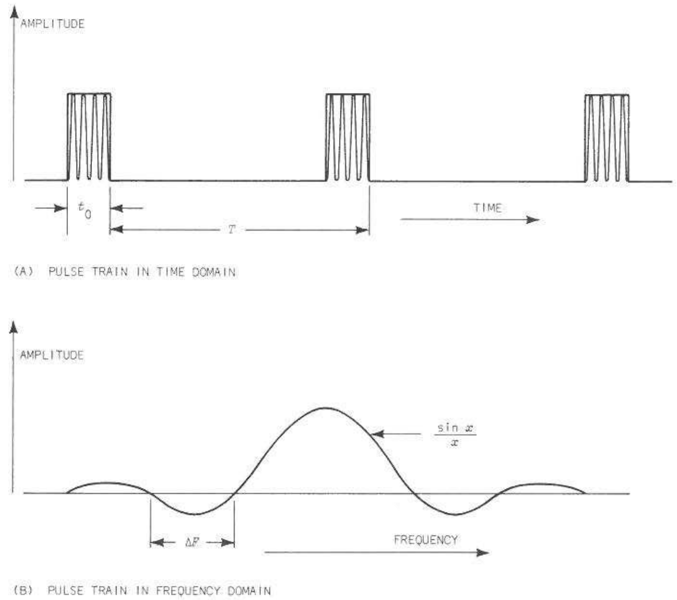


Fig. 5-6. Pulsed-signal analysis.

Another way of looking at the $\frac{\sin x}{x}$ shape is that this represents the energy distribution of a single pulse. Here the curve is not an envelope or locus curve, but is the actual shape of the spectral distribution. The spectrum is dense and continuous, and one no longer speaks of individual harmonics or CW responses. The detailed reasoning involved in these two interpretations will be found in Chapter 3. What is of interest here is the processes in the analyzer that lead to one or the other type of display.

Let us consider the dense spectrum first. Here the object is to obtain the spectral distribution of a single pulse. Unfortunately, this cannot be done with the type of spectrum analyzer under discussion. This is because the speed of relative frequency translation between the filter and signal is limited, so only a small range of the frequencies of interest can be checked during the short time that the pulse exists. One way out of this dilemma is to use many analyzers operating in parallel and look simultaneously at the different frequency portions of the same pulse. Another way is to have many identical pulses at which a single analyzer can look sequentially. Here many systems working a short time with one pulse have been traded for one system working a long time with many pulses. As long as the many pulses are identical and each succeeding measurement is independent of all previous measurements, the results of the two configurations will be identical. In considering dense spectra, we use the many-pulses, single-analyzer system. While the requirement that all pulses in the pulse train have the same characteristics can only be controlled by the source of the signal, the need for each successive measurement to be independent of all previous measurements is under the control of the spectrum-analyzer user. All that is needed to make each measurement independent of all others is that all traces (e.g., electron beam deflections) of the previous measurements be dissipated in the interval between measurements. Since measurements are performed one per pulse, the measurement interval is the interpulse interval. Also, the memory of a circuit is essentially proportional to its time constant, so the shorter the time constant the less trace is left from previous measurements in a given time interval. Finally, time constants are inversely proportional to bandwidths and interpulse intervals are inversely proportional to repetition rates. Hence, *for each measurement to be independent of all others, it is necessary that the resolution bandwidth be greater than the pulse-train repetition rate.* Let us now consider what actually happens in circuits.

one measurement per pulse

pulse train

Fig. 5-7 shows a time/frequency diagram similar to Fig. 5-3, only the input consists of a pulse train. Each of the pulses, theoretically, has a broad spectrum so the signal exists at many frequencies simultaneously. However, the signal is not continuous in time, occurring in narrow bursts, t_0 seconds in duration every T seconds. The distributed spectrum of each pulse is in turn moved in frequency by mixing with the sweeping local oscillator, resulting in the mixer output as shown in Fig. 5-7. Every intersection between the mixed sweeping spectrum and the stationary filter results in an output. As long as the pulse rate $(1/T)$ is less than filter bandwidth B, each input pulse produces an output which is independent of all the other pulses so the resultant display has the shape of the theoretical single-pulse dense spectrum.

$B > \frac{1}{T}$

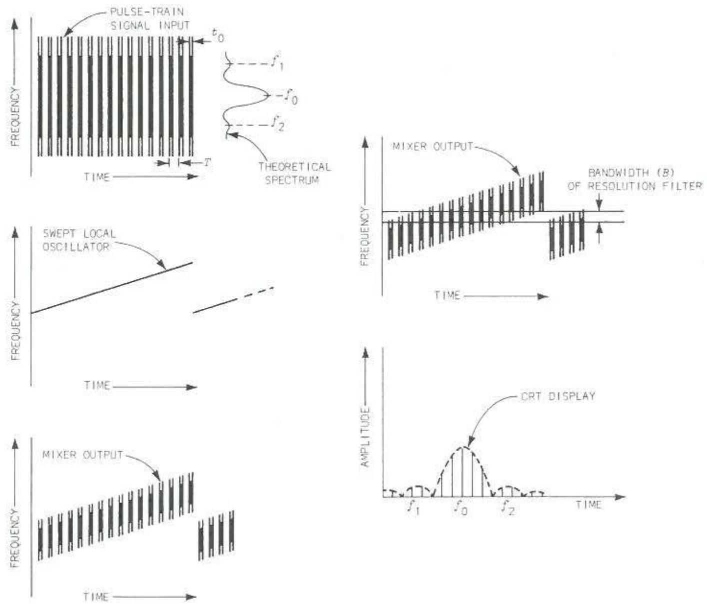


Fig. 5-7. Time-frequency diagram for pulsed-signal analysis.

Each time a pulse occurs there is an output. Hence, the composite output consists of lines, one line per pulse. This means that the total number of lines is equal to the number of pulses per sweep. As the sweep time is changed, the number of lines will change. These lines are not Fourier or spectral lines, but are strictly determined by the number of pulses per sweep and are usually termed *rep-rate lines*. Since the number of lines forming the spectrum depends on the sweep time, it is necessary that the sweep time be sufficiently long to generate enough pulse intercepts to define adequately the shape of the spectrum. It has been found experimentally that 10 lines per main lobe and 5 lines per side lobe is about the minimum that one can use.

B vs τ

Besides the requirement that resolution bandwidth be greater than the pulse repetition rate, there is also a constraint on the resolution bandwidth as a function of pulse width. The involvement with pulse width cannot be avoided, since, in truth, there are no CW signals, only short pulses and long pulses. We must answer the question of when a long pulse (seconds, minutes, hours or days in duration) can be treated as a CW signal. The matter can be considered from a time-domain or frequency-domain point of view, and both analyses lead to the same result.

CW is a long pulse

From a time-domain point of view, a long pulse might as well be a continuous wave if it exists long enough to trace the shape of the resolution filter. From Fig. 5-3 this means that the spectrum analyzer cannot distinguish between a CW signal and a pulse whose time duration is greater than $\tau = \frac{B}{D} T$, as given by equation (5-1). Offhand one might think that τ would be made as small as desired simply by decreasing the sweep time T. But there is an optimum beyond which one runs into trouble as given by equation (5-5). $T = \frac{D}{2.27B^2}$. Substituting for T, one gets:

$$\tau = \frac{B}{D} \cdot \frac{D}{2.27B^2} = \frac{1}{2.27B} . \qquad (5\text{-}7)$$

This means that, from a time-domain point of view, the demarcation line, between a pulse looking like a pulse or like a CW signal, is a pulse width about one-half of the inverse of resolution bandwidth. From a frequency-domain point of view, a long pulse looks like a CW signal when the resolution bandwidth is sufficiently wide to encompass most of the spectrum energy.

Most of the pulse energy is in the main lobe, which has a frequency width of $\frac{2}{t_0}$, where t_0 is pulse width. Hence, we come to the conclusion that $t_0 B \approx 2$ is the demarcation line between a pulse-like spectrum and a CW-like spectrum.

t_0 vs B for detail

Naturally, a resolution bandwidth which is just on the border line will not permit the display of the fine detail of a pulsed spectrum. It has been found experimentally that for adequate detail the pulse-width bandwidth product should be less than one-tenth, thus:

$$t_0 B \leqslant 0.1 \qquad (5\text{-}8)$$

A major difference between a pulse-type response and a CW-type response is in the width of the pulse that the final amplifier has to pass. In the CW case, we are dealing with the relatively wide, τ, pulse due to the steady-state response of the resolution amplifier, while, in the pulse case, we are dealing with the much narrower pulse, t_0, resulting in a transient response of the resolution amplifier. While the continuous-type spectrum is of major interest in pulsed RF, one can obtain either type of display by simply changing the resolution bandwidth of the spectrum analyzer. Table 5-1 details the major differences between the two types of display.

CW-TYPE SPECTRUM	CONTINUOUS-TYPE SPECTRUM
LINES ON SCREEN ARE FOURIER SPECTRAL COMPONENTS	LINES ON SCREEN ARE REPETITION-RATE LINES
LINE SPACING DEPENDS ON DISPERSION SETTING AND IS INDEPENDENT OF SWEEP TIME	LINE SPACING IS DETERMINED BY SWEEP TIME AND IS INDEPENDENT OF DISPERSION SETTING
MATHEMATICAL DESCRIPTION IS FOURIER SERIES	MATHEMATICAL DESCRIPTION IS FOURIER INTEGRAL
RESOLUTION SETTING IS B < REP RATE	RESOLUTION SETTING IS B > REP RATE
THE CRT DISPLAY SHOWS THE AMPLIFIER STEADY-STATE RESPONSE	THE CRT DISPLAY SHOWS THE AMPLIFIER TRANSIENT RESPONSE
THERE IS STILL ENERGY IN THE CIRCUIT FROM PREVIOUS PULSES WHEN THE NEXT PULSE OCCURS	ALL THE ENERGY IN THE CIRCUIT FROM THE PREVIOUS PULSE HAS DECAYED TO ZERO WHEN THE NEXT PULSE OCCURS
BANDWIDTH-PULSEWIDTH PRODUCT IS $Bt_0 > 1$	BANDWIDTH-PULSEWIDTH PRODUCT IS $Bt_0 < 1$

Table 5-1.

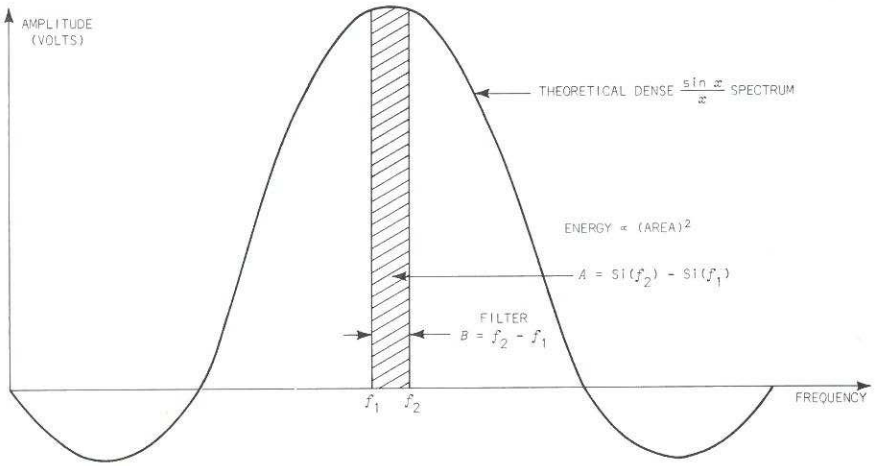


Fig. 5-8. Continuous-spectrum energy distribution.

SENSITIVITY

noise

Sensitivity is defined as a rating factor of the ability of the spectrum analyzer to display signals. Sensitivity is usually specified as the signal power which is equal to the analyzer noise power at a particular bandwidth; this is known as the "signal-plus-noise is equal to twice-noise" method. Spectrum analyzer noise level determines the sensitivity — less noise means better sensitivity. *All* amplifiers generate noise. Even an ideal amplifier would generate thermal noise, because of random current fluctuations in the input impedance. Thermal noise power can be computed from

$$N = kTB,$$

where,

k = Boltzmann's constant
= $1.37 \cdot 10^{-23}$ watt-seconds/degree,

T = absolute temperature
(measured from absolute zero, i.e., -273° Celsius – usually assumed to be 290° Absolute)

noise vs bandwidth

The thermal noise power and, hence, the sensitivity is directly proportional to the bandwidth. A wider bandwidth means a poorer sensitivity. For example, an ideal amplifier having a 1-MHz bandwidth has a theoretical sensitivity of -114 dBm, while the same amplifier with a 100-kHz bandwidth has theoretical sensitivity of -124 dBm. Such calculations lead to the conclusion that for best sensitivity the spectrum analyzer should be operated at narrow resolution bandwidths. The conclusion is correct for discrete CW signals but not for pulse signals. This is because a discrete signal has a spectrum which, at least in theory, has zero frequency width. A reduction in bandwidth reduces the noise but should have no effect on the signal. On the other hand, a pulse signal, which generates a continuous dense type of spectrum is affected by the resolution bandwidth. This is because the power level of a continuous spectrum is defined on a per-unit-bandwidth basis, as discussed in Chapter 3. This point is graphically illustrated in Fig. 5-8. Here is shown the continuous $\frac{\sin x}{x}$ spectral distribution, typical of a rectangular pulse; energy is proportional to area squared, so the wider the resolution

bandwidth the greater is the intercepted area and the larger the output. Eventually, when the bandwidth is so large that it intercepts most of the area, there is no further increase in output with an increase in bandwidth. However, the resolution bandwidth cannot be increased to this point without losing the fine details of the spectrum. As discussed previously, for proper definition of spectrum details the bandwidth should be about one-tenth of the inverse of the pulse width, $Bt_0 \leqslant 0.1$. Thus, for the same peak amplitude in time domain, a continuous spectrum for a pulsed signal will have a smaller amplitude than the discrete spectrum of a CW signal. This is intuitively apparent from the observation that equal peak amplitudes mean equal instantaneous power, which in one case is concentrated at one frequency and in the other case is distributed over a range of frequencies. Thus, there is a loss in sensitivity for pulsed signals as compared to CW signals.

less amplitude from pulsed signals

The formula from which the loss in pulsed-signal sensitivity can be computed was first reported in Volume XI of the *Radiation Laboratory Series*[5]. This was later shown to be a simplified version of a more complicated expression[6]. The simplified expression, which is sufficiently accurate for our purposes, is:

$$\alpha = \frac{3}{2} t_0 B$$

$$\alpha_{\mathrm{dB}} = 20 \log_{10} \frac{3}{2} t_0 B \qquad (5\text{-}9)$$

where t_0 is pulse width and B is 3-dB bandwidth. Fig. 5-9 is a graph of this relationship. Therefore, for discrete spectra the best sensitivity is obtained at narrow resolution bandwidths while, for continuous spectra, best sensitivity occurs at wide resolution bandwidths.

[5] Montgomery, "Technique of Microwave Measurements," *Radiation Laboratory Series*, Vol. XI, Sec. 7-2.

[6] Metcalf, et al., "Investigation of Spectrum Signature Instrumentation," *IEEE Trans*, EMC, June, 1965.

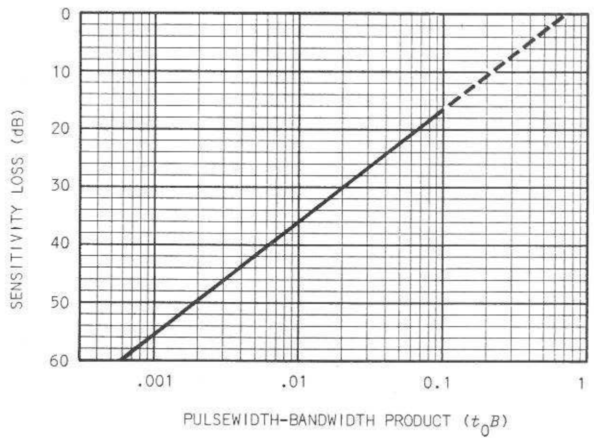


Fig. 5-9. Loss in sensitivity, pulsed-RF versus CW.

APPENDIX

CONVOLUTION

The convolution theorem is given in the Fourier transform section of Chapter 3. There it is indicated that the convolution of two time functions, $f_1(t)$ and $f_2(t)$, leads to a frequency-domain description which is the product of the two frequency-domain functions, $F_1(\omega)$ and $F_2(\omega)$:

$$\int_{-\infty}^{+\infty} f_1(\tau)\, f_2(t-\tau)\, d\tau \longleftrightarrow F_1(\omega)\, F_2(\omega) \qquad (5\text{-}10)$$

Convolution is of special interest because it is a mathematical description of the relative translation of two functions, $f_1(t)$ and $f_2(t)$, where the variable (τ) indicates the relative movement of the functions. Convolution, therefore, describes the process taking place in the sweeping-signal spectrum analyzer, where $f_2(\tau)$ is the stationary filter and $f_2(t-\tau)$ is the moving signal. Since the integral of a unit impulse is unity, it follows that convolution with an impulse leads back to the original function:

"folding"

$$\int_{-\infty}^{+\infty} f(\tau)\, \delta(t-\tau)\, d\tau = f(t) \qquad (5\text{-}11)$$

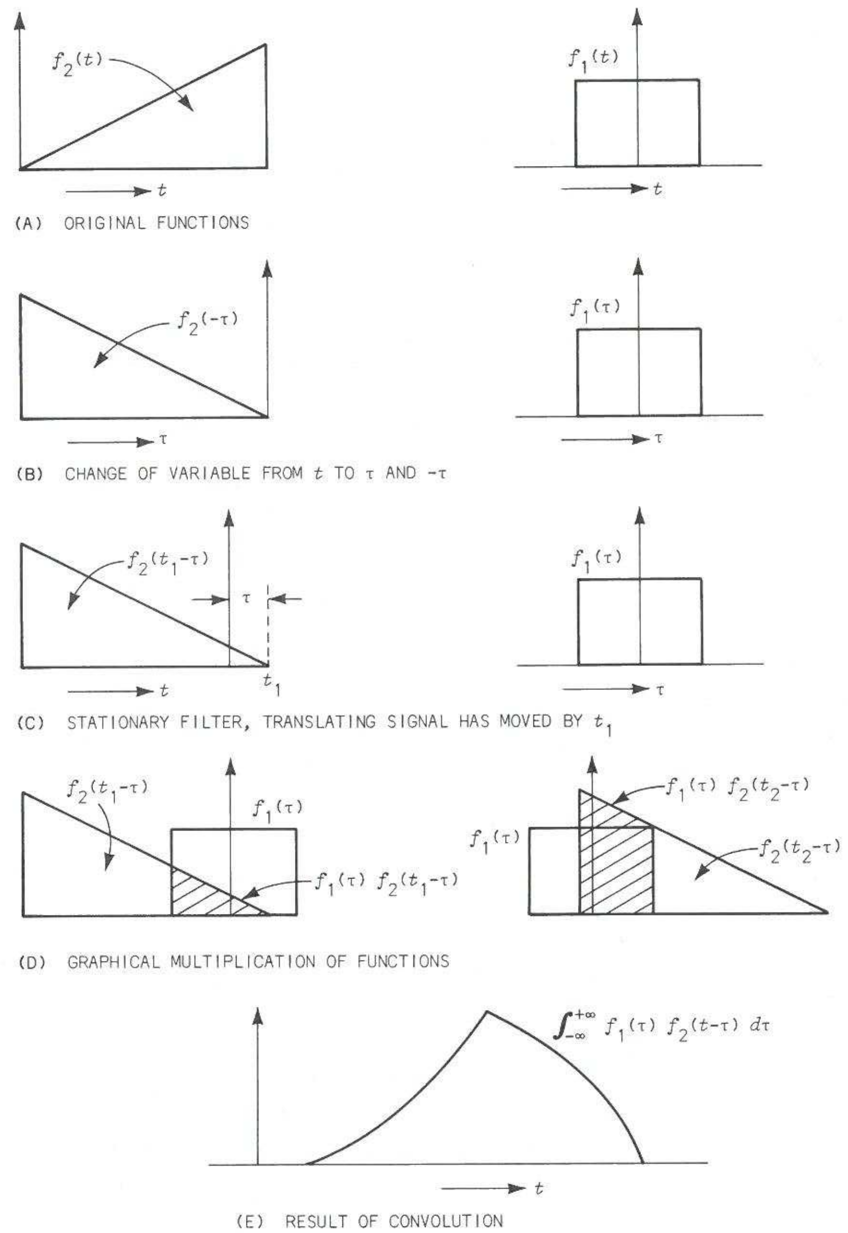


Fig. 5-10. Graphic illustration of convolution.

This means that the narrower the resolution bandwidth, the closer is the displayed spectrum to the theoretical input spectrum. Of course, there are constraints, such as sweep speed and dispersion, on how narrow the resolution bandwidth can be made. Nevertheless, for theoretically ideal spectrum analysis the resolution bandwidth has to be infinitely narrow.

The effect of the width of the sampling function, which in our case is determined by the resolution filter, is best illustrated graphically (see Fig. 5-10). We assume a triangular function $f_2(t)$ sliding past a stationary rectangular sampling function $f_1(t)$, as shown in Fig. 5-10A. We get the graphs in Fig. 5-10C which are the functions to be multiplied and integrated by the intermediate step, shown in Fig. 5-10B, which is just a change of variable.

Fig. 5-10D illustrates the actual integration, which is just a determination of the area under the multiplied function as shown by the shaded area. Fig. 5-10E is the final result, a distorted version of $f_2(t)$. As the width of the sampling function $f_1(t)$ gets narrower, the final result looks more and more like the input $f_2(t)$. In the limiting case, when the resolution filter has an infinitesimal bandwidth, the output function becomes identical to the input function. As a practical matter, it has been found that for a $\frac{\sin x}{x}$ spectral distribution, a filter bandwidth less than one-tenth of lobe width reproduces the original function with sufficient fidelity.

EXAMPLES

1) A spectrum analyzer is specified as having a certain sensitivity at a 100-kHz resolution bandwidth. The unit seems not to meet its specifications by 5 dB. The measurement conditions are:

Resolution bandwidth = 100 kHz

Dispersion = 1 MHz/div or 10 MHz full screen

Sweeptime = 10 μs/div or 100 μs full screen

What is wrong? From equation (5-3), the loss in sensitivity due to too fast a sweep rate is:

$$\alpha = \left[1 + 0.195\left(\frac{D}{TB^2}\right)^2\right]^{-\frac{1}{4}}$$

Substituting we have

$$\alpha = \left[1 + 0.195\left(\frac{10^7}{10^{-4} \cdot 10^{10}}\right)^2\right]^{-\frac{1}{4}} = \left(1 + 19.5\right)^{-\frac{1}{4}} = \frac{1}{2.13}$$

or a loss of $20 \log_{10} 2.13 \approx 6.6$ dB.

This accounts for the loss in sensitivity. Note that we cannot use precise numbers: Bandwidth B in equation (5-3) is the 3-dB bandwidth, while the resolution bandwidth is frequently defined as the 6-dB bandwidth. However, because of errors caused by differences between the actual and assumed phase responses, it is impossible to get an accurate number no matter what bandwidth is used. All that the above calculation can tell us is that the discrepancy of 5 dB is not unreasonable.

2) It is desired to observe the spectrum of an FM modulated signal. The approximate deviation is 100 kHz and the approximate FM rate is 5 kHz with a 100-MHz carrier. These numbers are well within the capability of many spectrum analyzers, including the Tektronix 491. However, the FM signal is initiated by an explosion and is expected to last no more than 1 ms. Is the measurement still possible?

We calculate the optimum resolution bandwidth which, from (5-5), is:

$$B_o = 0.66\sqrt{\frac{D}{T}}$$

$$B_o = 0.66\sqrt{\frac{10^5}{10^{-3}}} = 6.6 \text{ kHz.}$$

The actual resolution is $R_o = \sqrt{2}\, B_o \approx 9.3$ kHz.

Conclusion: The FM sidebands cannot be resolved.

3) It is desired to check a radar set operating at the rate of 10 pulses per second; what is the fastest reasonable sweep time for the analyzer? Assuming that we will observe the main lobe and one side lobe on either side, we need at least 20 rep-rate lines for appropriate definition of the spectrum envelope. Thus, we have to sweep at two seconds per screen diameter or slower.

4) The radar set in example 3 uses 1-μs pulses; what dispersion should we use? The width of a side lobe is $\frac{1}{t_0}$ and the width of the main lobe is $\frac{2}{t_0}$. To observe the main lobe and two side lobes we need a full screen dispersion of $\frac{4}{t_0} = 4$ MHz or 400 kHz/div.

5) For proper spectral envelope definition, what is the widest permitted resolution bandwidth? The relationship is $t_0 B \leqslant 0.1$, resulting in $B = \frac{0.1}{t_0} = 100$ kHz.

6) What is the loss in sensitivity compared to CW under these conditions? The formula is $\alpha = \frac{3}{2}\, t_0 B$; therefore,

$$\alpha = \frac{3}{2}\, 10^{-6} \cdot 10^5 = \frac{3}{2}\, 10^{-1}$$

$$\alpha_{dB} = 20 \log_{10} \frac{3}{2}\, 10^{-1} = -16.5 \text{ dB,}$$

or a loss in sensitivity of 16.5 dB.

7) Suppose the pulse rate is increased from 10 Hz to 200 kHz. Can the spectral distribution still be obtained? The answer is *NO*! We need a resolution bandwidth greater than the pulse repetition rate of 200 kHz in order to get a distributed spectrum display, but a resolution bandwidth which does not meet the requirements of $t_0 B \leqslant 0.1$ does not give adequate definition of the spectrum shape.

PART II

MEASUREMENT PRACTICE

6

THE MEASUREMENT PROBLEM

measurement need

There are four basic types of measurements that are desirable: Absolute frequency, relative frequency, absolute amplitude and relative amplitude. Not all spectrum analyzers have the sophisticated circuits which permit the absolute measurement of amplitude, however, most measurements do not require this. The most frequent need is for the determination of relative amplitude and relative frequency. Though there are, at most, four different types of measurements, some spectrum determinations can get quite complex. This is because in most instances it is not just a matter of making a single relative-amplitude measurement or a relative-frequency measurement. Most measurement problems call for a succession of several measurements, the sequence of which is important; these measurements may each call for different spectrum-analyzer control settings; and finally the measured data may have to be correlated or modified by computation before the final results are obtained. These aspects of measurements, which might be termed signal interpretation, are considered in subsequent chapters. In this chapter we will consider the four basic parameters:

Relative frequency
Absolute frequency
Relative amplitude
Absolute amplitude

FREQUENCY

Most spectrum analyzers have two frequency-related controls, CENTER FREQUENCY and DISPERSION. The CENTER FREQUENCY is an absolute-frequency control while the DISPERSION is a relative-frequency control. Of course, changes in the CENTER FREQUENCY setting give relative or frequency difference numbers. Thus, if a signal component is at the center of the CRT when the CENTER FREQUENCY dial reads 100 MHz and a different signal component tunes to the CRT center when the CENTER FREQUENCY dial reads 150 MHz, we conclude that the frequency difference between these signal components is 50 MHz. Here we have determined the relative frequency by taking the difference between two absolute-frequency readings. The problem with this method of relative-frequency measurement is poor accuracy. This is particularly true at high center frequencies. Thus, a 50-MHz frequency difference can be reasonably measured at a center frequency of 100 MHz but is difficult to determine accurately at a center frequency of 10 GHz. Because of these problems, it is recommended that, whenever possible, frequency differences be determined by using the DISPERSION rather than the CENTER FREQUENCY controls. When a frequency difference greater than the full-screen dispersion range is involved, the only way to make the measurement is to use the CENTER FREQUENCY control, otherwise the measurement is made by use of the DISPERSION control.

frequency measurement

The DISPERSION control is calibrated directly in terms of frequency difference. Thus, for the Tektronix Type 491, for example, the DISPERSION control is calibrated in a sequence of numbers from 1 kHz/div to 10 MHz/div. Suppose the operator wishes to determine the frequency difference between two signal components which appear on the CRT separated by 3.8 divisions and the dispersion setting is 5 MHz/div. The answer is 3.8 div · 5 MHz/div = 19 MHz. The accuracy of this method is independent of center frequency, permitting the determination of quite small frequency differences, such as 1 kHz, at very high frequencies, such as 40 GHz.

AMPLITUDE

Some of the newer spectrum analyzers have amplitude controls which are calibrated in terms of the absolute level of the spectral components. For example, the Tektronix Type 1L5 has a front-end attenuator that is calibrated for the deflection factor in units of RMS V/div. This permits the measurement of the absolute level of the various spectral components which comprise the input signal. Some spectrum analyzers are also calibrated in dBm when operating in the logarithmic vertical mode. In addition, all spectrum analyzers contain an attenuator, either at the input or further back in the system, which can be used for accurate relative-amplitude measurements. The measurement is performed by using the calibrated attenuator to reduce the display amplitude of the larger signal component to the former level of the smaller signal component. The amount of inserted attenuation is the ratio, or difference in dB, between the two signal components. Relative-amplitude measurements can also be performed by comparing the display height in divisions between the various signals with the spectrum analyzer in the linear vertical mode of operation. Direct measurement off the graticule is also possible in the logarithmic vertical mode, provided the graticule is calibrated in dB per division. Among these techniques, the most accurate is usually that involving the use of a calibrated attenuator, because amplifier nonlinearity does not affect the measurement.

measurement by attenuator

by graticule

MEASUREMENT LIMITATIONS

While the basic technique of frequency or amplitude measurement is simple, there are many points to be aware of when performing these measurements. These points pertain to generally good measurement practice and an awareness of the spectrum-analyzer limitations. Thus, a relative-amplitude difference of 100 dB cannot be measured with any presently available spectrum analyzer, a point which is easily determined from the maximum dynamic-range specifications. A more subtle point might be the fact that, although the specifications for the Tektronix Type 491 indicate a resolution bandwidth of less than 1 kHz and an on-screen dynamic range of at least 40 dB, it is impossible to observe a small signal 40 dB down from a larger signal removed by only 1 kHz. This is because the 1-kHz specification pertains to the 6-dB bandwidth, while the 40-dB bandwidth is much wider. Let

us, therefore, consider some of the spectrum-analyzer measurement limitations and some of the methods, or tricks of the trade, that can be used to improve a spectrum analyzer's measurement range.

ABSOLUTE FREQUENCY

All spectrum analyzers have a dial or other type of readout device that indicates the frequency that is supposed to correspond to the center of the CRT. Unfortunately, spectrum analyzers have spurious responses, so the readout does not always represent a true indication of an incoming signal. The major problem in absolute-frequency measurement is, therefore, to differentiate between the true response and the spurious responses. There are many types of spurious responses and these affect the spectrum analyzer to a different degree.

spurious response, 3 types

Three types of spurious responses affect the center-frequency readout capability. These are: *IF feedthrough*, *image* and *harmonic conversions*. These have different effects and are identified in a different manner, depending on whether the spectrum analyzer is swept front-end or of the swept IF variety. Let us consider the swept IF first.

Fig. 6-1 is a block diagram of a basic swept IF system. A portion of the input frequency spectrum, having a maximum frequency width equal to the bandwidth of the first amplifier, is translated in frequency and applied to the second mixer where it is treated the same as an input to a swept front-end system.

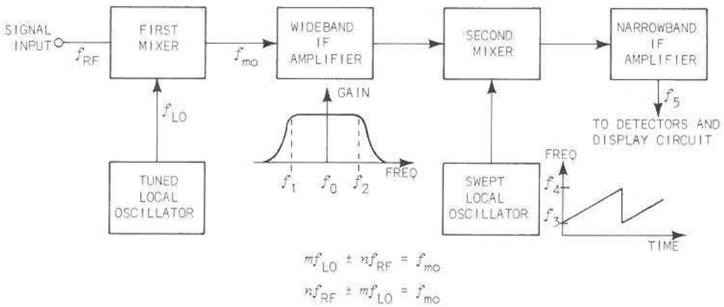


Fig. 6-1. Swept IF spectrum analyzer, basic block diagram.

In a properly designed system, the second conversion should not generate any spurious responses because the wideband amplifier controls the frequency width applied to the second mixer. Hence, our interest is in the first frequency conversion, which is described by equation (6-1).

$$mf_{LO} \pm nf_{RF} = f_{mo},$$
$$nf_{RF} \pm mf_{LO} = f_{mo}, \tag{6-1}$$

where:

f_{LO} = local oscillator frequency

f_{RF} = signal input frequency

f_{mo} = mixer output frequency

m, n = positive integers including zero

The CENTER FREQUENCY dial, or readout, is normally calibrated to indicate the local-oscillator frequency f_{LO}, or a harmonic of f_{LO}, in combination with the IF-amplifier center frequency f_0. From equation (6-1) the possible combinations are:

$$f_d = mf_{LO} + f_0,$$
$$f_d = mf_{LO} - f_0, \tag{6-2}$$
$$f_d = f_0 - mf_{LO}.$$

When the mixer output frequency (f_{mo}) is equal to the IF center frequency (f_0), there is a spectrum-analyzer response. The dial frequency (f_d) may or may not be the same as the signal input frequency (f_{RF}) at that time. If the dial and signal frequencies are the same, we are dealing with a true response, otherwise the response is spurious.

IF feedthrough

Consider, for example, the IF feedthrough. This occurs when the input signal frequency is equal to the IF-amplifier center frequency, thus $f_{RF} = f_0$. The local-oscillator frequency (f_{LO}) has nothing to do with this response, hence from (6-2), it is clear that the dial indication (f_d) has no validity for this response. The IF-feedthrough spurious response can only occur for the narrow range of input frequencies that fall within the passband of the first IF amplifier. This frequency range is indicated as f_1 to f_2 in Fig. 6-1. The IF-feedthrough spurious response is recognized by the fact that the setting of the RF-frequency tuning dial has no effect on it.

image

Another bothersome spurious response is the image. The image occurs when the signal frequency satisfies one of the mixer equations, but the dial is calibrated for one of the other two possible conversions. For example, suppose the dial is calibrated for $f_{LO} - f_d = f_0$. An input signal of frequency $f_{RF} = f_d$ would satisfy this equation and be a true response. However, an input signal whose frequency satisfies the equation $f_i - f_{LO} = f_0$, where $f_i \neq f_d$, will also appear on the screen. This second response is the image. The image frequency (f_i) and that of the true response, corresponding to the dial setting (f_d), are separated by twice the IF center frequency (f_0). This can be shown as follows:

$$f_{LO} - f_d = f_0 \quad \text{true response}$$

$$-f_{LO} + f_i = f_0 \quad \text{image response}$$

$$\text{sum,} \quad f_i - f_d = 2f_0 \tag{6-3}$$

In a swept IF system, the image is recognized by the fact that the signal display will move across the screen in the opposite direction to the true response. This will be recognized from equation (6-1), where for one conversion the output frequency *in*creases with increasing local-oscillator frequency, while for the other conversion the output frequency *de*creases with increasing local-oscillator frequency. This is illustrated below.

$$\uparrow f_{LO} - f_d = f_{mo}\uparrow,$$

$$f_i - \uparrow f_{LO} = f_{mo}\downarrow$$

where $\uparrow$ means increasing frequency and $\downarrow$ means decreasing frequency.

Of course, in order to determine which of the two responses is the image, one has to know which of the three possible conversions corresponds to the dial calibration. When it is discovered that the on-screen response is the image, it is necessary that the dial setting be changed by twice the wideband IF center frequency in order to obtain the true response. Whether the dial numbers have to be increased or decreased depends on which conversion the dial is calibrated for.

The last of the spurious responses that can cause ambiguity in absolute-frequency measurements is the harmonic conversion response. This is due to the signal combining with a harmonic of the local oscillator to produce an IF frequency output at a dial setting which does not correspond to the signal frequency.

harmonic conversions

Not all harmonic conversions are spurious responses. Many spectrum analyzers utilize harmonic conversions to increase frequency coverage. Thus, in the Tektronix Type 491, the fundamental, second, third, fifth, and tenth local-oscillator harmonic conversions are used. The other harmonic conversions such as fourth or sixth are considered spurious responses. Harmonic spurious responses are identified by the rate of movement across the CRT as a function of RF center-frequency tuning. This is because the rate at which the mixer output frequency (f_{mo}) changes is determined by $m_t f_{LO}$, while the number on the dial changes at the rate of $m_d f_{LO}$. Unless the two harmonic numbers, m_d and m_t, are the same, the rate of frequency change on the dial will not agree with the rate of frequency change on the CRT. In practice, the test is to make a frequency difference measurement using the RF center-frequency dial and to compare the numbers obtained with a frequency difference determined from the dispersion. If the

dial change vs frequency change

response on the screen is a harmonic spurious response, the two numbers will disagree by an integer fraction such as 1/2, 2/3, 4/3, etc. If, for example, the dial is calibrated for m_d = 1, while the signal frequency is such that to get the IF center frequency (f_0) it must combine with m_t = 2, tuning the signal across the full dispersion of the screen will result in a dial number change which is one-half that of the dispersion.

SWEPT FRONT END

So far, we have considered how to verify the center-frequency reading for a swept IF system. Similar problems exist, though not to the same extent, for the swept front-end system. Fig. 6-2 is a basic block diagram of a swept front-end unit. The spurious response problem is much alleviated in this system because the first IF amplifier, being of the narrowband variety, can be constructed at a much higher frequency than the wideband unit in a swept IF system. A higher IF amplifier frequency means greater frequency separation between the true response and the spurious responses and, hence, fewer difficulties. Of course, one

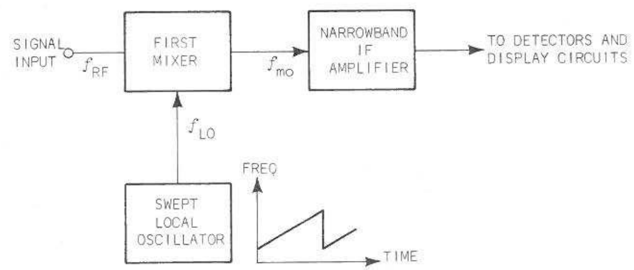


Fig. 6-2. Swept front-end spectrum analyzer, basic block diagram.

pays for this with greater sweeper-system complexity — hence, more weight, size and cost. While this type of system has the same type of spurious responses as the swept IF unit, the appearance and identification of these responses are different and sometimes more difficult. Let us consider these spurious responses in turn.

The IF feedthrough response is caused by an input signal whose frequency is within the passband of the first IF amplifier, and which does not enter into a conversion or mixing with the first local oscillator. In the swept front-end unit, the first local oscillator is the sweeping local oscillator. Hence, the IF feedthrough is not a swept signal. This means that a continuous-wave input is not converted into its frequency-domain equivalent of a narrow pulse. The continuous-wave input exists at all times within the passband of the IF amplifier, causing the whole baseline to deflect or rise. The IF feedthrough spurious response is, therefore, recognized by a shift in the baseline level.

As in the swept-IF system, the image frequency is separated from the true response frequency by twice the IF amplifier center frequency. However, unlike the swept-IF system, the image response does not move on the screen in an opposite direction to the true response as the center-frequency dial is tuned. These responses will move in opposite directions with input-signal frequency change but will behave the same with local-oscillator frequency change. This point is demonstrated in the time/frequency diagram, Fig. 6-3. Two CW signals are applied to the spectrum analyzer at frequencies f_{RF1} and f_{RF2}, respectively. One uses the $f_{LO} - f_{RF1} = f_{mo}$ response, while the other corresponds to the $f_{RF2} - f_{LO} = f_{mo}$ response. One of these responses is the image, while the other is the true response – for our purposes it does not matter which is which. As the sweeping local-oscillator center frequency is changed, the frequency sawtooth moves from the curve labeled f_{LO1} to the one labeled f_{LO2}.

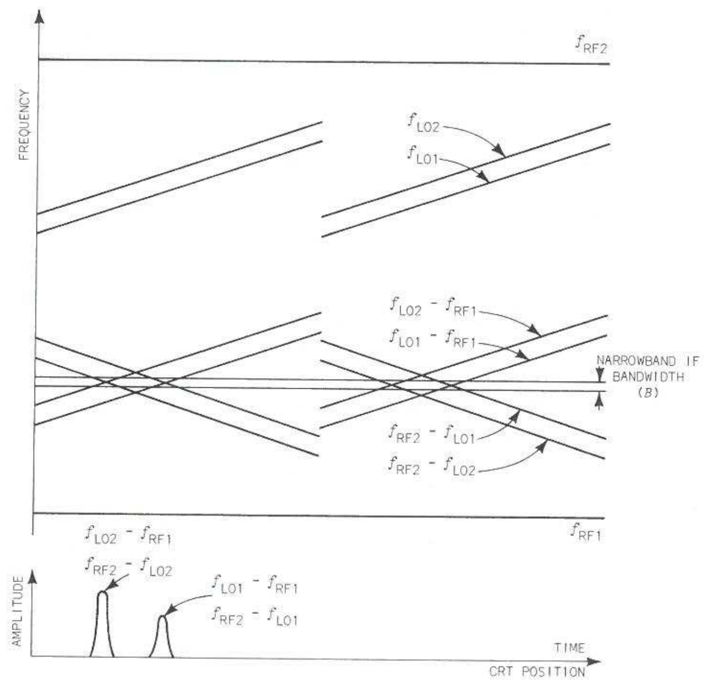


Fig. 6-3. Time/frequency diagram of image response tuning in swept front-end spectrum analyzer.

The corresponding mixer outputs are $f_{LO1} - f_{RF1}$ and $f_{LO2} - f_{RF1}$ for one response and $f_{RF2} - f_{LO1}$ and $f_{RF2} - f_{LO2}$ for the other response. From Fig. 6-3 it will be observed that regardless of which conversion is utilized the response on the CRT screen moves the same way. Hence, there is no way to distinguish the image from the true response simply by operating the spectum-analyzer controls in a swept front-end system. The only way to observe a different effect on the CRT display for these two responses is to change the frequency of the actual input signal.

In the case of harmonic conversions, the swept front end behaves the same as the swept IF system. These spurious responses are identified in both systems by noting differences in signal tuning rate as compared to RF center-frequency dial indications.

AMPLITUDE MEASUREMENTS

Amplitude measurements are of two types – relative and absolute. All spectrum analyzers can be used for relative-amplitude measurements, but only some have the capability of absolute-amplitude measurements without resorting to external signal generators for amplitude calibration. While many spectrum-analyzer parameters, such as conversion flatness, gain stability, etc., help determine whether the unit is capable of absolute-amplitude measurements, the user need not be aware of the details. The user can determine whether his instrument is or is not capable of absolute-amplitude measurements by simply referring to the specifications. The establishment of the instrument limitations for relative-amplitude measurements is, however, not a simple matter. This aspect of amplitude measurements will now be considered.

Relative-amplitude measurements are of three types:

I. Measurement of the relative amplitude between a large signal and a nonharmonically related small signal.

II. Measurement of harmonic distortion, which means finding the relative amplitude between a large signal and a harmonically related small signal.

relative amplitude

III. Intermodulation distortion, which means the measurement of the relative amplitude between small and large signals in the presence of more than one large signal.

I. Nonharmonically Related Small Signal In The Presence Of Large Signal

The spectrum analyzer has an on-screen dynamic-range specification which implies that one can measure the amplitude difference between a small signal and a large signal up to the difference specified. This is not strictly correct. There are many situations where a larger amplitude difference can be measured. There are also situations where the specified range cannot be achieved. These ramifications are best illustrated for a specific case. For example, the Tektronix Type 491 Spectrum Analyzer is specified to have an on-screen dynamic range of 40 dB in the LOG vertical mode. First we will consider the limitations on the measurement and then the technique for making larger than 40-dB difference measurements.

A) Limitations on amplitude difference measurements.

There are three limitations on amplitude difference measurements. These are:

maximum useful dynamic range

1) The *maximum useful dynamic range* is defined as the amplitude difference between the analyzer sensitivity and the maximum input power that the analyzer can accommodate in a linear fashion. The sensitivity is specified for each instrument as a function of both frequency and resolution bandwidth. The maximum input power is also specified for each instrument, and is -30 dBm for the Type 491. Thus, for the Type 491, for example, in the 275-to-900-MHz frequency range at 100-kHz resolution, the sensitivity is -90 dBm and the maximum useful dynamic range is 90 - 30 = 60 dB. The maximum input power can sometimes be increased to -20 dBm, getting a 10-dB improvement in dynamic range. This, however, depends on many parameters and each case must be considered separately. When in doubt, go by the specification which is -30 dBm.

skirt of response curve

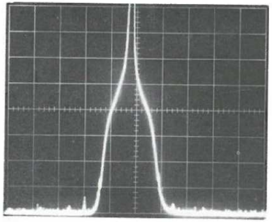
2) *Skirt selectivity* is a measure of resolution bandwidth at more than 6 dB down. The skirt selectivity limits how close, in frequency, the small signal can be to the large signal and still be resolved as a separate signal. This is a function of the variable-resolution amplifier and is different for different units. Using the Tektronix Type 491 as our example, the skirt-selectivity shape is illustrated in Fig. 6-4, where one division above the baseline is 60 dB down, dispersion is 100 kHz/div and resolution bandwidth is about 3 kHz at 6 dB down. The 60-dB down bandwidth is about 250 kHz, indicating that a small signal, 60 dB down from a large signal, has to be separated by at least 125 kHz to be resolvable. Cutting the resolution bandwidth improves the sensitivity and hence the maximum useful dynamic range. However, this has very little effect on the limitation of the frequency separation between the large and small signal.

3) *Analyzer gain suppression* – when the spectrum-analyzer circuits are overdriven by a large signal, the gain becomes suppressed or reduced in the vicinity of the large signal. A small signal in the gain-suppressed region will also be suppressed and will not show up on the analyzer. Hence, besides the skirt-selectivity requirement, it is necessary that the small signal be sufficiently separated from the large signal in frequency, so as not to be in the gain-suppressed region.

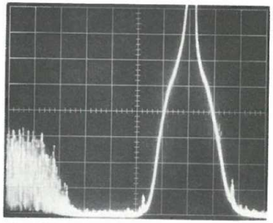
gain suppression

The amplitude of the large signal has a major effect on the frequency width of the gain-suppressed region. The larger the signal the worse the suppression. Hence, it is advantageous to operate at the narrowest resolution, therefore, best sensitivity and, therefore, lowest signal level possible.

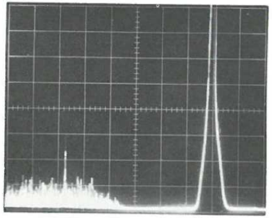
Since signal level is the major parameter in establishing the suppression characteristics, it is important that the analyzer gain be set as low as possible commensurate with optimum sensitivity. This means that the gain should be set for less than half of a division of noise. Any increase in gain will degrade the suppression characteristics. Fig. 6-4 illustrates the above points:



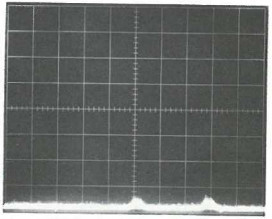
(A) VERTICAL DISPLAY - LOG
DISPERSION - 100kHz/DIV
RESOLUTION - 3kHz
FIRST DIVISION BOTTOM - 60dB DOWN



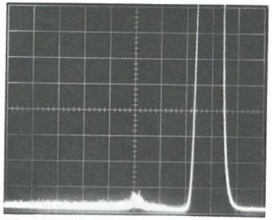
(B) SAME AS (A) EXCEPT GAIN IS INCREASED



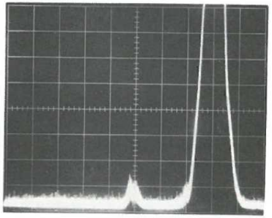
(C) VERTICAL DISPLAY - LIN
DISPERSION - 200kHz/DIV
RESOLUTION - 3kHz
FIRST DIVISION BOTTOM - 60dB DOWN



(D) VERTICAL DISPLAY - LIN
DISPERSION - 200kHz/DIV
SIGNAL LEVEL AT -83dBm



(E) VERTICAL DISPLAY - LIN
DISPERSION - 200kHz/DIV
LARGE SIGNAL AT -23dBm
SMALL SIGNAL AT -83dBm



(F) SAME AS (E) EXCEPT LOG MODE

Fig. 6-4. Skirt-selectivity and gain suppression shapes.

Fig. 6-4A shows a small signal 60 dB down with respect to a large signal. Signal separation is 200 kHz. The separation characteristics are determined by skirt selectivity only, no gain suppression is observed. Good gain suppression characteristics were obtained by operating at 3-kHz resolution, permitting a -97 dBm small signal to be easily observed and the large signal at -37 dBm is well below the -30 dBm input limit. Further, the vertical mode is LOG and the gain is set just high enough to see some noise.

Fig. 6-4B shows the effect of increased analyzer gain. All parameters are the same as for Fig. 6-4A but the gain has been increased so that the noise level has increased by about a factor of ten. Note that the gain is suppressed for about 300 kHz from the skirt of the large signal. The small signal situated about 50 kHz from the skirt of the large signal is barely recognizable.

Fig. 6-4C shows the effect of operating in the LIN mode. Here we have a gain-suppressed region about 800 kHz wide. The small signal has been moved to 1.2 MHz from the large signal.

Fig. 6-4D shows two -83 dBm signals at about 600 kHz separation in LIN mode.

Fig. 6-4E has one of the signals of Fig. 6-4D increased in amplitude by 60 dB to -23 dBm. There is gain suppression for about 500 kHz. However, for signal separations greater than 500 kHz, we have no difficulty in making a 60-dB dynamic range measurement, in spite of the fact that the large signal is -23 dBm, 7 dB greater than the specified limit of -30 dBm.

Fig. 6-4F settings are the same as Fig. 6-4E, except that the vertical is in LOG mode. Note that there is no gain suppression.

B) Measurement technique for large dynamic range measurements.

The major limitations in relative-amplitude measurements are sensitivity and signal separation. The ultimate limit on signal separation is skirt selectivity, so, unless the signals are sufficiently separated, even measurements within the specified dynamic range may not be possible.

measurement procedure

The recommended measurement procedure follows. Adjust the resolution and gain for about a quarter to one-half of a division of noise in the LOG mode. Using attenuators, not the gain control, reduce the level of the input signal so the small signal is about 1 division high. Care should also be taken to maintain the large signal below about -30 dBm. Now insert either external or IF attenuators to reduce the level of the large signal to that of the former level of the small signal. Sometimes there is not sufficient attenuator range to make the measurement. For example, in the Tektronix Type 491, the IF attenuator has a 51-dB range. If the amplitude difference is less than 51 dB, then the IF attenuators will suffice. If the difference is greater than 51 dB, proceed as follows: Insert 51 dB of IF attenuation and observe the level of the large signal (e.g., 3.5 div). Remove 10 dB or some other amount of IF attenuation; the signal will get larger. Now reduce the signal to its former level, 3.5 div in our example, by means of the gain control. The signal level has now been reduced by 51 dB and there is still some IF attenuation left to reduce the signal further to the former level of the small signal.

II. Harmonic Distortion

Again using the Tektronix Type 491 as an example, the maximum dispersion is 100 MHz. This means that, except for signals at less than 50 MHz, harmonics cannot be observed together with the fundamental. At higher frequencies it is quite likely that the second harmonic will fall within a different dial scale than the fundamental.

Since large changes in sensitivity occur from scale to scale, it is generally impractical to make reasonably accurate harmonic-level measurements without using a signal generator for calibration purposes.

A second problem is the harmonic-distortion characteristics of the spectrum analyzer itself. The smaller the input level, the better the distortion properties of the analyzer.

There are two ways of determining whether the observed harmonics are part of the signal or generated by the analyzer:

filter the input

A) By means of filters – If the distortion is in the signal, then a low-pass filter should eliminate the harmonics from the analyzer CRT. Likewise, a high-pass filter will eliminate the fundamental but the harmonics should still be displayed. Similar reasoning applies to bandpass and band-reject filters.

amplitude ratio vs gain change

B) By amplitude effects – If appropriate filters are not available, it is still possible to tell where the distortion is coming from. The method is based on the fact that the distortion generated by the analyzer occurs only when a circuit is operating in a nonlinear mode. Thus, the ratio of fundamental to harmonic is a function of input level. Therefore, if the analyzer has sufficient dynamic range at the frequency in question, the input level should be either increased or decreased (say by 6 dB) and the harmonic level relative to the fundamental remeasured. If there is no change, then the harmonics are part of the original signal; if there is a small change (e.g., less than a dB), then most of the measured distortion is part of the signal. If there is a substantial change (e.g., 3 dB), then the distortion is coming from the analyzer and the measurement is not valid.

III. Intermodulation

Intermodulation distortion (IM) occurs when two or more large input signals mix with each other to produce additional signals not in the original input. These intermodulation products appear as additional signals separated in frequency from the original signals by the frequency separation of the original signals. As with other nonlinear effects, one can determine whether the intermodulation is due to the spectrum analyzer or is part of the signal by repeating the measurement at a different signal level. This technique is described in detail in the previous section on harmonic distortion.

The size of the IM products depends on the amplitude level and frequency separation of the input signals. Instrument performance gets worse as the signal level is increased and signal separation is decreased. Again using the Tektronix Type 491 as a specific example, the effect of separation is most noticeable at signal separations between 50 kHz and 500 kHz. Very little improvement is obtained by separating the signals more than 500 kHz and very little degradation is observed when the separation is decreased to less than 50 kHz.

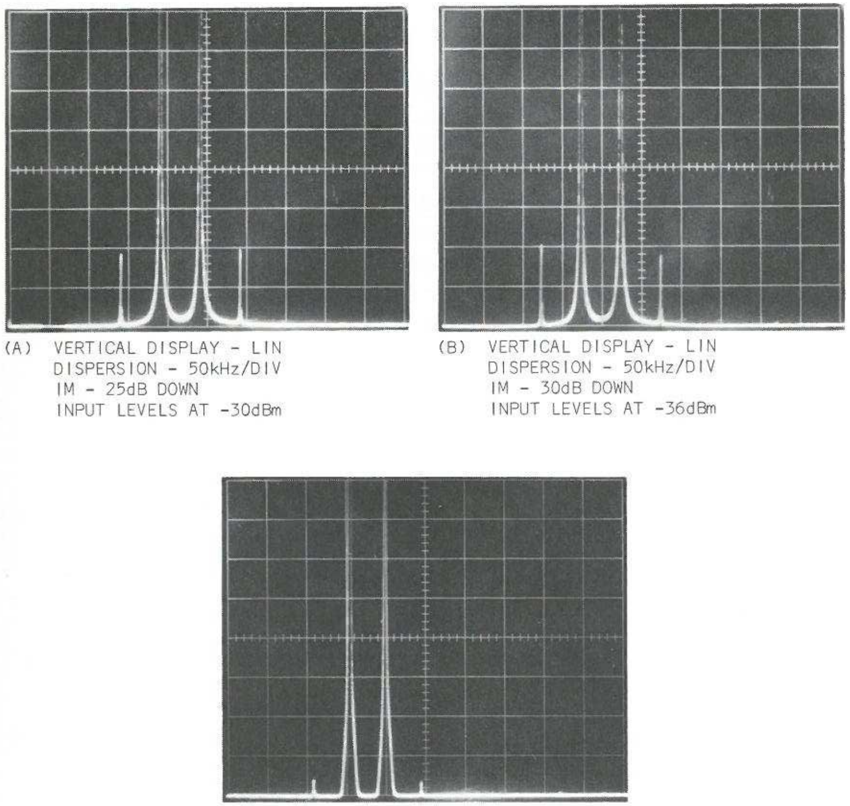
The ultimate limit on signal separation is determined by skirt selectivity, since the two signals must be separated (should not blend into each other) at the baseline. The input levels should be kept at less than -30 dBm, since performance degrades drastically at larger inputs.

The photographs in Fig. 6-5 illustrate the above.

Fig. 6-5A shows two signals at -30 dBm each, separated 50 kHz. The vertical is linear and the IM products are 25 dB down.

Fig. 6-5B illustrates the effect of signal level on IM. The input levels have been decreased to -36 dBm. The IM products are now 30 dB below the input.

Fig. 6-5C shows the effect of signal separation. Here the signals are separated by 500 kHz and the IM is 40 dB down.



(A) VERTICAL DISPLAY - LIN
DISPERSION - 50kHz/DIV
IM - 25dB DOWN
INPUT LEVELS AT -30dBm

(B) VERTICAL DISPLAY - LIN
DISPERSION - 50kHz/DIV
IM - 30dB DOWN
INPUT LEVELS AT -36dBm

(C) VERTICAL DISPLAY - LIN
DISPERSION - 500kHz/DIV
IM - 40dB DOWN
SIGNAL LEVELS AT -30dBm

Fig. 6-5.

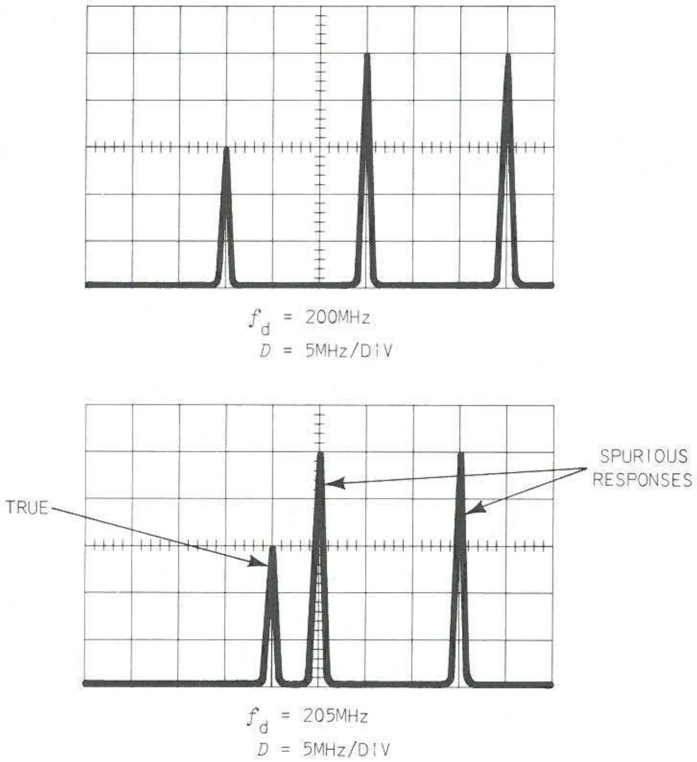


Fig. 6-6. Identification of image response.

SIGNAL FREQUENCY IDENTIFICATION PROBLEMS

1) Refer to Fig. 6-6 which represents the CRT display of a swept-IF spectrum analyzer at two dial settings, 200 MHz and 205 MHz; the dispersion is 5 MHz/div. Which of the responses are true and which are spurious? Clearly there is at least one spurious response, since, of the three responses, one moved to the right while two moved to the left as the center frequency was changed from 200 MHz to 205 MHz. However, before we can tell which of the responses is the true one, it is necessary to know for which mixer-conversion equation the dial is calibrated. For the Tektronix microwave swept-IF spectrum analyzers (1L20, 491, etc.), the true response is due to $f_{LO} - f_{RF} = f_{mo}$. Hence, as the dial is tuned to higher numbers, which means the local-oscillator frequency is increased, the output frequency should increase and the true signal should move from left to right.

An image response would at the same time move from right to left. This analysis leads to the identification of the signals as shown in Fig. 6-6.

Besides identifying the true response, we can also identify the frequency of the signal causing this response. At a dispersion of 5 MHz/div, we would have to tune the dial five more megahertz, or to 210 MHz, in order to get the true response in the center of the screen, so the true response is caused by a 210-MHz signal. The image-response frequencies cannot be determined without a knowledge of the IF center frequency. Assuming that this is 200 MHz, such as for the Tektronix Type 491, we reason as follows: One image signal is at the center of the screen when the dial reads 205 MHz; hence, the local-oscillator frequency is computed from $f_{LO} - f_d = f_0$, or f_{LO} = 205 + 200 = 405 MHz. The image conversion is $f_i - f_{LO} = f_0$, or f_i = 405 + 200 = 605 MHz. At a dispersion of 5 MHz/div, the other image signal is at a frequency 15 MHz above the first one, or 620 MHz.

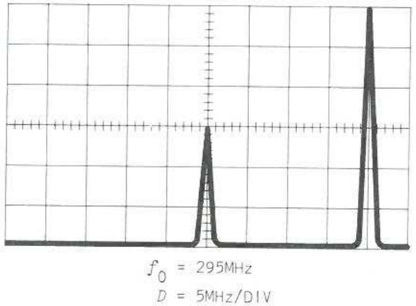


Fig. 6-7. Frequency of several true responses.

2) Refer to Fig. 6-7; assume that you have determined that both responses are true. Without changing the center-frequency setting, determine the input frequencies. The smaller of the two responses is at the center of the screen, hence, its frequency corresponds to the center-frequency dial setting which is 295 MHz. The second signal is removed from the first one by four graticule divisions and, at a dispersion of 5 MHz/div, this corresponds to a frequency difference of 20 MHz. However, do we add this 20 MHz to 295 MHz or do we subtract it? The answer is that for Tektronix instruments we subtract it. The reasoning is as follows: The true conversion equation is $f_{LO} - f_{RF} = f_{mo}$. Hence, a larger input frequency (f_{RF}) means a smaller mixer output frequency (f_{mo}) and vice versa. Since it is the mixer output that is ultimately displayed, we must go by what f_{mo} does rather than what f_{RF} does, even though it is f_{RF} that we wish to identify. The screen corresponds to low frequency on the left and high frequency on the right, but the frequency we are talking about is f_{mo}. At a fixed local-oscillator setting, a larger f_{mo} means a smaller f_{RF}; hence, we subtract 20 MHz from 295 MHz to get 275 MHz as the second frequency. Note that as far as the input signal f_{RF} is concerned, *in a swept-IF system, using the local-oscillator frequency above the signal frequency conversion, the higher input frequency signal is on the left.*

3) Refer to Fig. 6-8 which shows a CRT display involving five signals. Which of the responses are true, which are spurious and what are the input frequencies?

finding true response

For a swept-IF system with local-oscillator frequency above signal frequency, a true response will move from left to right as the dial frequency is increased. Response *a* is the only one that moves from left to right with increasing dial frequency, all the other responses either move from right to left or stand still. Hence, *a* is either a true response or a harmonic conversion and all the other responses are spurious. To determine whether *a* is a true response, we note that, at a dispersion of 5 MHz/div, *a* has moved four divisions as the center frequency dial has moved from 300 MHz to 320 MHz. Since 320 - 300 = 4 · 5, the dispersion and the tuning dial agree; therefore, *a* is a true response. The input frequency causing this response is 305 MHz because: When the center of the CRT corresponds to 300 MHz, *a* is one division, or 5 MHz, to the left of center – for a converted signal, left means higher frequency, hence, 305 MHz.

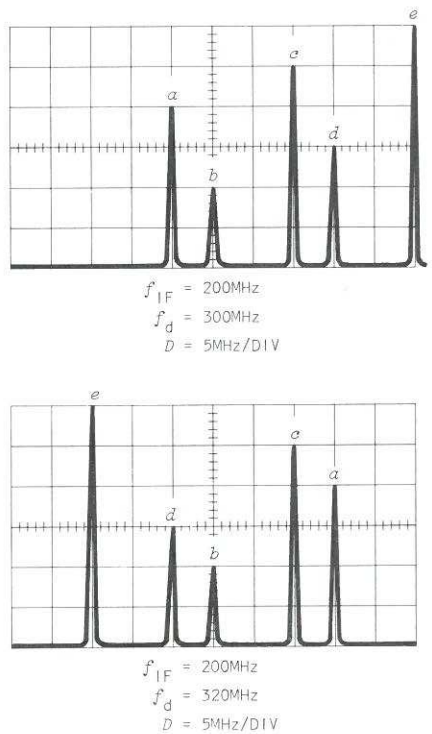


Fig. 6-8. Complex spurious display.

Let us now take the spurious responses in alphabetical order. Response *b* is in the center of the screen and has not moved. It is an IF-amplifier feedthrough response, which makes it 200 MHz. Response *c* is also an IF feedthrough because it has not moved. This response is two divisions away from center so the signal is 2 · 5 = 10 MHz from 200 MHz. This signal is at 210 MHz because, for an unconverted signal, the screen represents increasing frequency going left to right.

Consider now response *d*. This response moves in the wrong direction, so it is an image. Response *d* moves 20 MHz when the dial moves 20 MHz, so it represents a prime conversion. Hence, when the dial reads 300 MHz, the local-oscillator frequency is $f_{LO} = f_d + f_{IF}$ = 300 + 200 = 500 MHz. For the image, $f_{RF} - f_{LO} = f_{IF}$; f_{RF} = 200 + 500 = 700 MHz for the center of the screen. But, when the dial reads 300 MHz, response *d* is three divisions to the right of center. While, for a true response, the screen represents increasing frequency left to right; for the image, it is right to left. Hence, response *d* is 3 · 5 = 15 MHz above screen center, or 700 + 15 = 715 MHz.

second harmonic image

Finally, response *e* is an image since it moves from right to left with increasing dial reading. This image is, however, not a fundamental conversion but a second harmonic conversion, since the response moves eight divisions, representing 40 MHz according to the dispersion, when the dial has moved only 20 MHz. Again, when the dial is 300 MHz, the local oscillator is 500 MHz. The second harmonic-image spurious response is based on the conversion $f_{RF} - 2f_{LO} = f_{IF}$. Hence, f_{RF} = 200 + 2 · 500 = 1200 MHz for the center of the screen. The response is, however, five divisions to the right of center, which for an image means higher in frequency. Hence, at 5 MHz/div, response *e* is due to a signal at a frequency of 1200 + 5 · 5 = 1225 MHz.

7

AMPLITUDE MODULATION

The following fundamental relationships apply to normal double-sideband amplitude-modulation (AM) measurements:

1) In the time domain, the percent modulation is computed from:

$$m = \frac{K - 1}{K + 1}; \qquad K = \frac{E_{max}}{E_{min}}, \tag{7-1}$$

E_{max} and E_{min} as in Fig. 7-1.

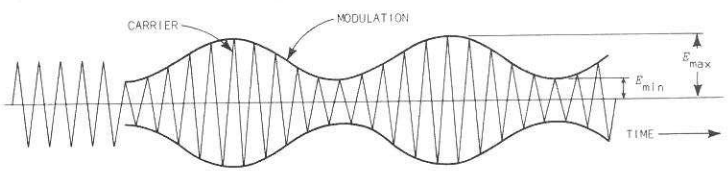


Fig. 7-1. Time-domain AM.

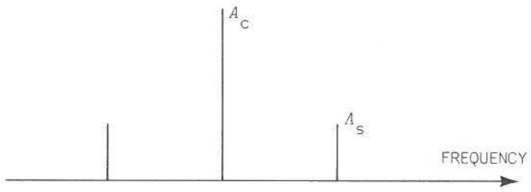


Fig. 7-2. Frequency-domain AM.

2) In the frequency domain, the percent modulation is computed from:

$$m = \frac{2A_s}{A_c}, \tag{7-2}$$

where A_s and A_c are per Fig. 7-2; $\frac{A_s}{A_c}$ is a voltage ratio.

3) The spectrum consists of a carrier and two equal amplitude sidebands for each modulating frequency. The sideband spacing with respect to the carrier frequency is equal to the modulating frequency.

4) In AM the carrier spectral component is of constant amplitude regardless of the degree of modulation.

The following figures illustrate normal AM measurements.

Fig. 7-3A shows the time-domain (oscilloscope) appearance of a 10-MHz carrier, modulated by a 10-kHz signal. At a sweep time of 50 μs/cm the period of the modulating wave is 100 μs, or a frequency of 10 kHz. The 10-MHz carrier frequency could also be determined by operating at a faster sweep. To determine the percentage modulation, we observe that

$$K = \frac{E_{max}}{E_{min}} = \frac{2}{0.8} = 2.5$$

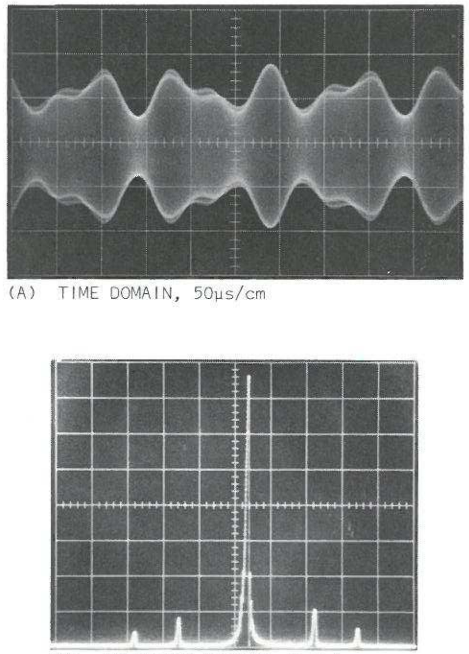
and

$$m = \frac{K-1}{K+1} = \frac{1.5}{3.5} \cong 43\%.$$

Fig. 7-3B shows the same signal in the frequency domain (spectrum analyzer). From the center-frequency setting we observe that the carrier is at 10 MHz. The modulating frequency is 10 kHz, since the sidebands are spaced 2 cm from the carrier at a dispersion of 5 kHz/cm. The percentage modulation is

$$m = \frac{2A_s}{A_c} = \frac{2}{4.8} \cong 42\%.$$

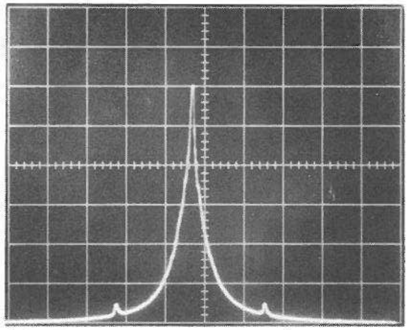
Besides yielding the same data as the oscilloscope, the spectrum analyzer also shows that the modulating signal has some second harmonic distortion as indicated by the additional set of small sidebands 20 kHz from the carrier.



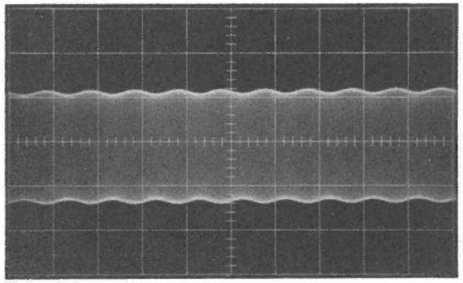
(A) TIME DOMAIN, 50μs/cm

(B) FREQUENCY DOMAIN, 5kHz/cm, 10MHz CF, LIN

Fig. 7-3. Single-tone AM.



(A) FREQUENCY DOMAIN, 5kHz/cm, 10MHz CF, LOG. SIDEBANDS ARE 30dB DOWN.



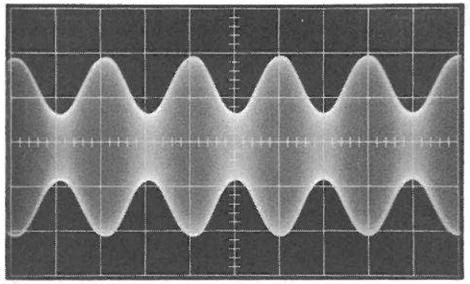
(B) TIME DOMAIN, 50µs/cm

Fig. 7-4. Low-level single-tone AM.

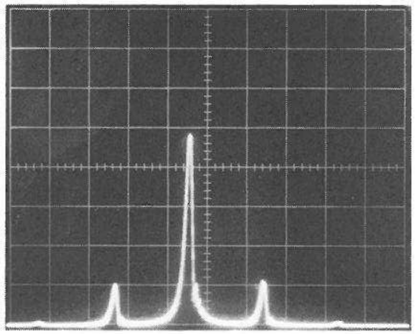
For relatively high levels of modulation (e.g., over 10%), the oscilloscope and spectrum analyzer yield basically the same information. However, for small levels of modulation the spectrum analyzer is definitely superior, as illustrated in Fig. 7-4. Both measurements indicate the frequencies, but the spectrum-analyzer determination of percentage modulation is much easier. Here the sidebands are 30 dB down, obtained by reducing the carrier to the former level of the sidebands by means of the spectrum analyzer's internal attenuator. Since 30 dB represent a voltage ratio of 0.0317, the percentage modulation is about 6%.

multitone modulation

Another case where data is easier to obtain from the spectrum is that of multitone modulation. This is illustrated in Fig. 7-5. Fig. 7-5A is the time-domain appearance of a multitone AM wave. Not only is it difficult to ascertain the degree of modulation, but it is virtually impossible to determine the frequencies involved. Fig. 7-5B is the frequency-domain appearance of the same waveform. From this it is apparent that there are two modulating frequencies, one 10 kHz and the other about 16 kHz. The percentage modulation at 10 kHz is about $2/7.8 \cong 26\%$ and, at 16 kHz, it is about $2(0.5)/7.8 \cong 13\%$. The reason for the unequal amplitude between the lower and upper sidebands is incidental frequency modulation. Combined AM and FM is discussed in more detail in Chapter 8.



(A) TIME DOMAIN, 50µs/cm



(B) FREQUENCY DOMAIN, 5kHz/cm, 10MHz CF, LIN

Fig. 7-5. Multitone AM.

OTHER FORMS OF AM

suppressed carrier or sideband

A form of AM that saves power is double-sideband suppressed-carrier modulation. Here the interest centers on the degree of carrier suppression rather than on the degree of modulation. The amplitude of the carrier is measured relative to the sidebands, usually with the transmitter operating at the rated peak envelope power (PEP). Fig. 7-6 is a spectrum analyzer display of a double-sideband suppressed-carrier amplitude-modulated wave. The carrier amplitude is one-sixth of each sideband; or in dB the carrier-to-sideband ratio is $20 \log_{10} 1/6 \cong -15.5$ dB. The time-domain appearance of double-sideband suppressed-carrier AM is similar to that of standard AM at 100% modulation. With some care these can, however, be distinguished. Fig. 7-7 illustrates the difference in the time-domain appearance of these two forms of modulation.

Another frequently utilized form of AM is single sideband. Here only one sideband is transmitted, while the other sideband and the carrier are suppressed. This saves power and conserves frequency space. The sideband suppression is measured in the same manner as the carrier suppression ratio previously discussed.

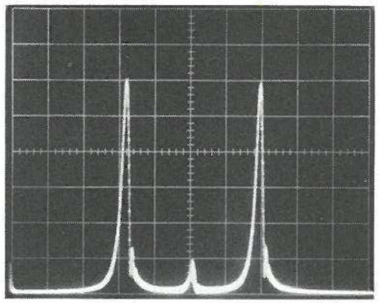


Fig. 7-6. Double-sideband suppressed-carrier AM in frequency domain. Vertical display is LIN.

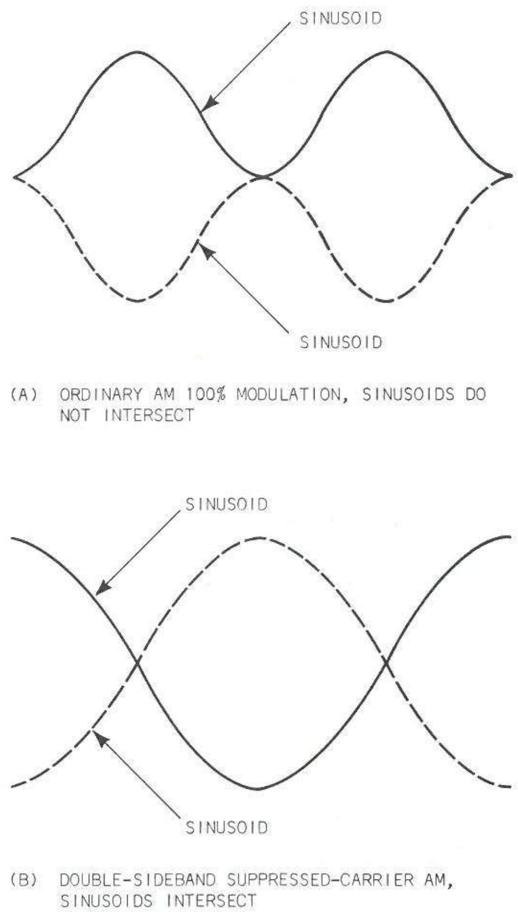


Fig. 7-7. Time-domain difference between ordinary 100% AM and double-sideband suppressed-carrier AM.

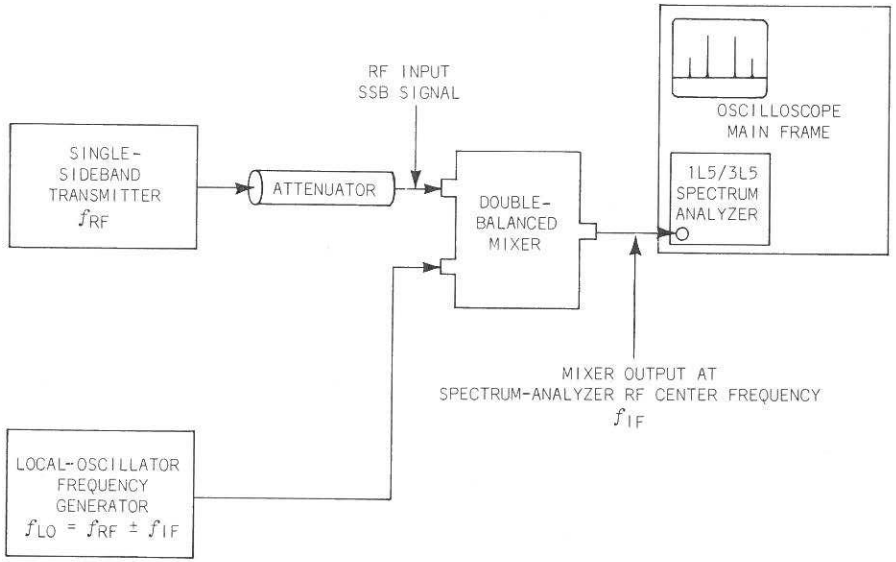


Fig. 7-8. Method of measurement of intermodulation distortion using Tektronix Type 1L5 or 3L5 Spectrum Analyzer, external local-oscillator signal generator and double-balanced mixer.

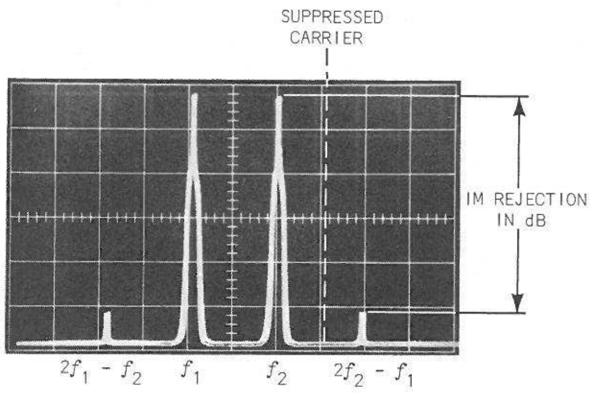


Fig. 7-9. Measurement of intermodulation rejection.

inter-
modulation
distortion

A measurement which is of particular interest in single sideband is that of *intermodulation distortion.* The measurement is performed by modulating the single-sideband transmitter with two audio tones and checking the output with a spectrum analyzer for additional sidebands. Frequently, the small frequency separation between the audio tones (e.g., 1 kHz and 1.6 kHz) and the high degree of intermodulation suppression that it is desired to measure (e.g., 50 dB) preclude the use of standard radio-frequency spectrum analyzers. One, then, has the choice of going to a specialized single-sideband test set or of heterodyning the signal down in frequency so that a standard audio-frequency spectrum analyzer, such as the Tektronix Type 1L5 or 3L5, may be used. Fig. 7-8 shows such a test setup. When the heterodyne method is used, it is important that the signal source, used as the local oscillator, be highly stable. A frequency synthesizer makes an excellent signal source for this application. Fig. 7-9 is a frequency-domain display of such a measurement.

8

FREQUENCY MODULATION

The following basic relationships apply to frequency modulation (FM).

1) Frequency modulation is a constant-energy process. The total energy of the modulated wave does not change as the degree of modulation changes.

2) The frequency-domain representation of an FM wave consists of a carrier and sidebands spaced in frequency around the carrier. The spacing between frequency components is equal to the modulating frequency (f).

3) Theoretically, the FM wave contains an infinite number of sidebands. The sideband energy, however, falls off very rapidly outside the peak frequency deviation. Deviation is measured with respect to the carrier frequency.

4) The amplitudes of the various frequency components, including the carrier component, change as the deviation changes. This is a consequence of the requirement that the total energy remain constant regardless of the deviation.

5) The relative amplitudes of the frequency components are in the same relationship as the relative amplitudes of Bessel functions of the first kind. Bessel functions of the first kind are designated by the letter "J." The complete characterization of the frequency component amplitudes is

$$J_p\left(\frac{\Delta F}{f}\right);$$

where p is called the order and represents the frequency component number ($p = 0$ for the carrier, $p = 1$ for the first sideband, etc.), and $\frac{\Delta F}{f}$ is called the argument and represents the modulation index. The modulation index

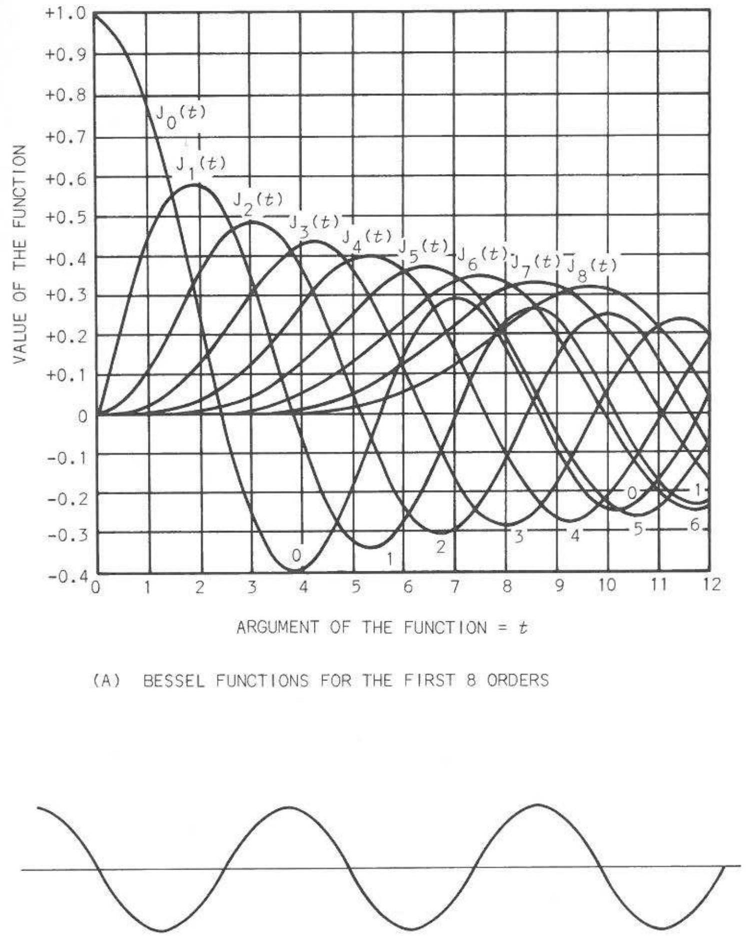
(sometimes designated as m, t, or β) is defined as the ratio: peak frequency deviation ΔF divided by the modulating frequency f.

6) Bessel functions are the solution to a certain differential equation, just as the standard trigonometric functions (sine and cosine) are the solution to a specific differential equation. Graphs and the tables of Bessel functions of the first kind are readily available. Fig. 8-1 is such a graph.

7) The information of interest in FM is: the carrier frequency (F), the modulating frequency (f), and the deviation (ΔF). The carrier frequency F is obtained by reading the spectrum-analyzer center-frequency dial and the modulating frequency f is obtained by calculating the frequency spacing between two adjacent components by use of the calibrated dispersion. The deviation (ΔF) can, however, not be determined directly. First, one obtains the modulation index from which the deviation is then calculated. Most of what follows pertains to the calculation of the deviation. A detailed theoretical discussion of FM will be found in Chapter 4.

Methods of determining the deviation differ, depending on the modulation index. Techniques that work well at fractional indices (e.g., $\Delta F/f \ll 1$) will not yield any useful data for relatively large indices (e.g., $\Delta F/f > 10$). While no standard designation exists, it is convenient, for the purposes of this book, to separate FM into three deviation regions. We shall call these narrowband, wideband and ultrawideband FM. It is emphasized that these are not standard designations and should not be confused with similar names found elsewhere. For the purposes of our discussion, narrowband FM means a modulation index less than unity, wideband FM will refer to modulation indices from about one to about ten, and ultrawideband FM will refer to modulation indices greater than ten.

Let us consider each of these in turn.



(A) BESSEL FUNCTIONS FOR THE FIRST 8 ORDERS

(B) cos ωt

Fig. 8-1.

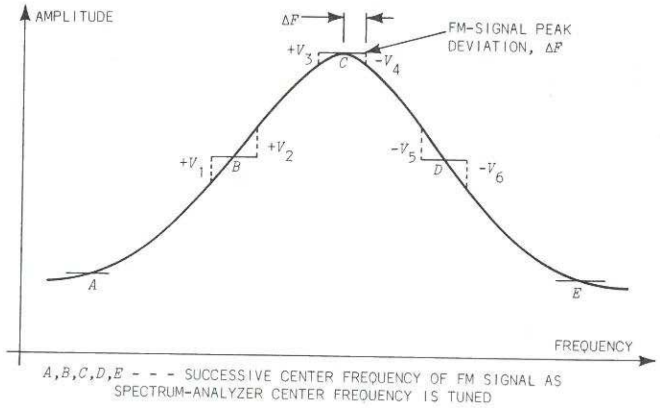


Fig. 8-2. Slope detecting an FM signal.

NARROWBAND FM

Because sideband energy falls off very rapidly outside the peak frequency deviation, narrowband FM is characterized by only two significant sidebands. This is especially true at modulation indices less than about 0.5, where it is difficult to distinguish, from a spectrum-analyzer display, whether the signal is AM or FM. If one knows that the signal is FM, one can proceed directly to the problem of determining the various modulation parameters. If, however, the basic nature of the signal is not known, it is necessary first to determine whether it is AM or FM. With spectrum analyzers that have a zero dispersion position, it is possible to do this for modulation rates up to about one-half the widest resolution bandwidth of the spectrum analyzer.

slope detection

AM is distinguished from FM by the basic difference in the methods used in detecting them. Amplitude modulation can be detected by an ordinary diode peak detector, whereas to detect frequency modulation it is necessary to use a discriminator. While spectrum analyzers do not usually contain a discriminator for detecting FM, this can be done by slope detecting the FM signal on the skirts or slopes of the resolution amplifier curve. This is illustrated in Fig. 8-2. The heavy horizontal lines represent the successive center frequencies of the frequency-modulated signal as it is tuned through the range of the resolution amplifier resonance curve by means of the spectrum-analyzer fine center-frequency control. With the spectrum analyzer set for zero-hertz-per-division dispersion (i.e., not sweeping), the detected modulating signal appears directly on the CRT. The amplitude of this detected signal depends on the slope of the resonance curve and the deviation of the frequency-modulated signal. At positions *A* and *E* the slope of the resonance curve is small, resulting in very little signal output. At position *B* the output voltage is $(+V_1) + (+V_2) = (V_1 + V_2)$. At position *D* the output voltage is $(-V_5) + (-V_6) = -(V_5 + V_6)$, where the minus sign denotes the change in slope between positions *B* and *D*. At position *C* the output is $V_3 - V_4$, which is very small if the curve is reasonably symmetrical. The result is that for an FM signal we have two positions of maximum output voltage occurring around the middle of the resonance curve. The edges and peak of the resonance curve yield very little output. Fig. 8-3 shows the actual spectrum-analyzer display for such a measurement. For amplitude modulation, there is only one position of maximum output – at the peak of the resonance curve. Having established, from prior knowledge or through the above procedure, that the signal in question is narrowband FM, we can now proceed with the basic measurement.

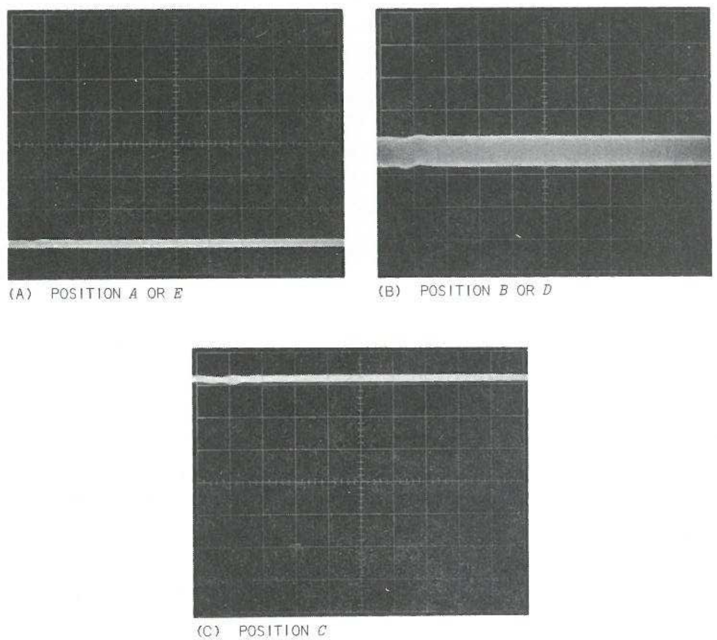


Fig. 8-3. Detecting FM with a Spectrum Analyzer as illustrated in Fig. 8-2.

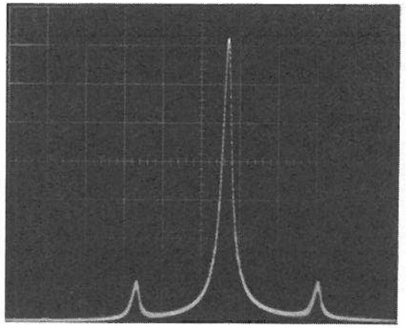


Fig. 8-4. Narrowband FM in frequency domain.

Fig. 8-4 is a frequency-domain representation of a narrowband FM signal. The carrier frequency of 275 MHz is determined from the spectrum-analyzer center-frequency dial. The modulating frequency is equal to the frequency spacing between the carrier and either of the first sidebands. At a dispersion of 20 kHz/cm and a sideband spacing of 2.5 cm, the modulating frequency is 50 kHz. To determine the deviation ΔF, we utilize the fact that the amplitude of the carrier is represented by the magnitude of the Bessel function of the first kind, zero order, with modulation index equal to argument; while the amplitude of either of the first sidebands is represented by the magnitude of the Bessel function of the first kind, first order and modulation index equal to the argument. Mathematically,

$$A_c \propto J_0\left(\frac{\Delta F}{f}\right)$$

$$A_{s1} \propto J_1\left(\frac{\Delta F}{f}\right)$$

In order to make an actual calculation, we need an equality rather than a proportional relationship. This is obtained by taking the ratio between the two amplitudes, thus:

$$\frac{A_{s1}}{A_c} = \frac{J_1\left(\frac{\Delta F}{f}\right)}{J_0\left(\frac{\Delta F}{f}\right)}, \tag{8-1}$$

where the ratio is of voltages rather than powers. The easiest way to determine the deviation is now to use the fact that for small modulation ratios the following relationship holds:

$$\frac{J_1\left(\frac{\Delta F}{f}\right)}{J_0\left(\frac{\Delta F}{f}\right)} = \frac{\Delta F}{2f}. \tag{8-2}$$

A variation of this is given in equation 4-25. From Fig. 8-4, the voltage ratio of the amplitudes is 0.9/7.2 = 0.125. Hence,

$$\frac{J_1\left(\frac{\Delta F}{f}\right)}{J_0\left(\frac{\Delta F}{f}\right)} = \frac{\Delta F}{2f} = 0.125,$$

and the modulation index is

$$\frac{\Delta F}{f} = 0.25.$$

Since the modulation frequency f = 50 kHz, the deviation ΔF = 0.25(50) = 12.5 kHz. If one does not recall the formula, or if it is desired to try for a more accurate result because the formula becomes inaccurate above modulation indices of about 0.5, the procedure is to use tables of Bessel functions. Here,

the first step is to let the $J_0\left(\frac{\Delta F}{f}\right)$ term equal unity, an obviously reasonable assumption for small values of argument, as demonstrated by the graph of Fig. 8-1. One then finds in the table the value of $J_1\left(\frac{\Delta F}{f}\right)$, which is equal to the measured ratio of sideband to carrier. With this as a first approximation, the calculation can be refined by checking for a closer match of the $\left(\frac{\Delta F}{f}\right)$ value in the vicinity of the first approximation. For example, suppose the voltage ratio of first sideband amplitude to carrier amplitude, as determined from the spectrum-analyzer display, is 0.375[1]. Assuming that $J_0\left(\frac{\Delta F}{f}\right) \cong 1$, we have $\frac{\Delta F}{f} \cong 0.8$, as illustrated in the partial table of Bessel functions, Fig. 8-5. Checking actual ratios of

$$\frac{J_1\left(\frac{\Delta F}{f}\right)}{J_0\left(\frac{\Delta F}{f}\right)},$$

in the vicinity of $\frac{\Delta F}{f} = 0.8$, we find that at $\frac{\Delta F}{f} = 0.7$,

$$\frac{J_1(0.7)}{J_0(0.7)} = \frac{.329}{.881} \cong 0.373.$$

The modulation index is, therefore, much closer to 0.7 than to 0.8.

[1]These numbers were chosen to illustrate the method. A ratio measurement to an accuracy of three significant figures is beyond the accuracy of most spectrum analyzers.

$\frac{\Delta F}{f}$	0.6	0.7	0.8	0.9
$J_0\left(\frac{\Delta F}{f}\right)$	0.912	0.881	0.846	0.808
$J_1\left(\frac{\Delta F}{f}\right)$	0.287	0.329	0.369	0.406

0.375

Fig. 8-5. Partial table of Bessel functions.

WIDEBAND FM

Most FM measurements are for modulation indices of about one to ten, which for purposes of this discussion has been designated wideband. As for small modulation indices, the method of measurement depends on the determination of ratios. This creates difficulties because in many instances it is important to determine the deviation to a very high degree of accuracy, whereas the performance of most spectrum analyzers precludes the measurement of relative amplitude to better than two significant figures. There is only one relative-amplitude determination which can be obtained to a high degree of accuracy; this is where one of the components is zero. The technique where one of the amplitude components, usually the carrier, is made to go to zero is known as the *carrier-null method* or the *Crosby null method* – after Murray G. Crosby who did much of the basic work on FM measurements.

carrier-null method

The carrier-null method of FM deviation measurement is the most used and also the most accurate of all methods. However, it is only applicable in cases where the FM signal can be changed during the measurement. Where the signal to be measured is fixed, it is necessary to use other, more complicated and less accurate, measurement techniques.

Most people have no problem with the fact that for a sinusoid, such as cos θ, there are specific values of the angle θ, where the magnitude of the sinusoid goes to zero. For cos θ this occurs when the angle theta is an odd multiple of $\pi/2$, that is, $\theta = 90^\circ, 270^\circ$, etc. A similar relationship occurs for Bessel functions, where the magnitude of the function goes to zero at certain specific values of modulation index. These points of zero amplitude are called *nulls* and, when referring specifically to the carrier component, the term *carrier null* is used. Carrier nulls occur at those modulation indices where the zero-order Bessel function of the first kind goes through zero. These points of zero carrier amplitude can be seen in Fig. 8-1, where the $J_0(t)$ curve crosses the zero axis.

Fig. 8-6 is a table of the first ten modulation indices at which the carrier goes through a null. The actual measurement procedure will now be illustrated by means of examples.

CARRIER NULL	MODULATION INDEX $\left(\frac{\Delta F}{f}\right)$
FIRST	2.4048
SECOND	5.5201
THIRD	8.6537
FOURTH	11.7915
FIFTH	14.9309
SIXTH	18.0711
SEVENTH	21.2116
EIGHTH	24.3525
NINTH	27.4935
TENTH	30.6346

Fig. 8-6. Table of carrier nulls.

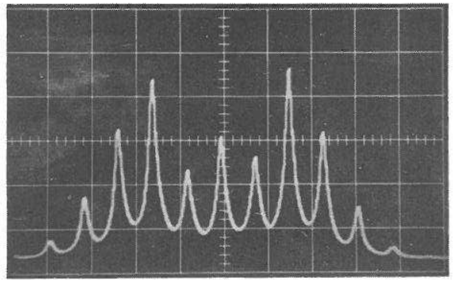


Fig. 8-7. FM deviation measurement.

EXAMPLE 1

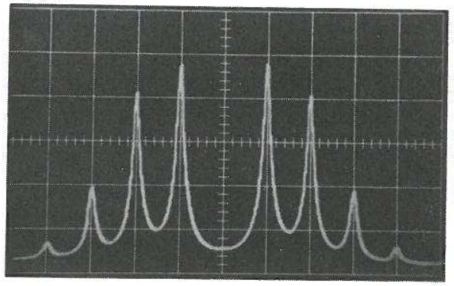
Fig. 8-7 shows a spectrum-analyzer display of a wideband FM signal. The sidebands are three-quarters of a graticule division apart and, at 10 kHz/div, this corresponds to a modulation frequency f of 7.5 kHz. The object is to determine the deviation. This will be accomplished by changing the modulation frequency f such that the modulation index $\frac{\Delta F}{f}$ corresponds to one of the carrier nulls. The deviation will then be computed from the known values of f and $\frac{\Delta F}{f}$. One of the dangers in this procedure is mistaking one carrier null for another. Thus, one might think that the display corresponds to the first carrier null at a modulation index of 2.4, whereas the actual modulation index is 5.5, corresponding to the second carrier null. The following procedure will guard against such an error.

carrier disappears at the null

First we guess at the possible limits of the deviation. From Fig. 8-7 it is observed that the sideband amplitudes start falling off at about 30 kHz from the carrier, which is at the center of the display; and there are virtually no sidebands beyond 40 kHz from the carrier. Our guess, therefore, is that the modulation index is probably around $\frac{\Delta F}{f} = \frac{30}{7.5} = 4$, but it might be as high as $\frac{\Delta F}{f} = \frac{40}{7.5} = 5.33$. Both our best and maximum-limit guess place us between the first and second carrier nulls which are at a modulation index of 2.4 and 5.5 respectively. If we are correct, we should be able to get to the first carrier null by increasing the modulating frequency f while maintaining the modulating signal amplitude constant[2]. As the modulating frequency is increased, the carrier amplitude is seen to decrease until it goes to zero, as shown in Fig. 8-8. In our example this occurs at a modulating frequency of 10 kHz. Since the first carrier null occurs at a modulation index of 2.4, we compute the deviation as:

$$\frac{\Delta F}{f} = 2.4,$$

$$\Delta F = 2.4 \cdot 10 = 24 \text{ kHz}.$$



DISPERSION - 10kHz/cm
VERTICAL - LIN

Fig. 8-8. Deviation measurement illustrating carrier null.

[2] For a linear modulator, the deviation is dependent only on the amplitude of the modulating signal.

Just to make sure that there is no mistake, we could also decrease the modulating frequency and, so, check the deviation at the second carrier null. The two results should of course agree, otherwise there is an error in the measurement. This is the most accurate method of determining an unknown deviation, since the modulation frequency can be measured to a very high degree of accuracy with a counter.

EXAMPLE 2

vary amplitude

Another method of getting a carrier null is to change the amplitude of the modulating signal source. The deviation is directly proportional to the modulating voltage amplitude when operating within the linear range of the modulator. Voltages, however, cannot be measured as accurately as frequencies, so this method is less accurate than that based on frequency measurement.

Frequently the object is not to determine what the deviation is but, rather, to establish a particular deviation. Such a case might be found in the broadcast industry where the Federal Communication Commission specifies 200-kHz channel separation with a maximum frequency deviation of 75 kHz with respect to the carrier and at a maximum modulating frequency of 15 kHz. It would be of interest to adjust such a transmission system so that the modulating voltage would not exceed the level corresponding to a 75-kHz deviation. The simplest way of establishing the level of the maximum voltage not to be exceeded, is to set the modulating frequency such that a deviation of 75-kHz will correspond to a carrier null; and then adjust for the null by changing the modulating voltage amplitude. Thus, at a deviation of 75 kHz, the first carrier null corresponds to a modulating frequency of

$$f = \frac{75}{2.4} \cong 31.2 \text{ kHz}.$$

This is greater than the maximum permitted modulating frequency of 15 kHz, so the first carrier null cannot be used here. The second carrier null results in a computed modulating frequency of $\frac{75}{5.5201} = 13.586$ kHz, which can be used. The sequence of photographs in Fig. 8-9 illustrates the procedure.

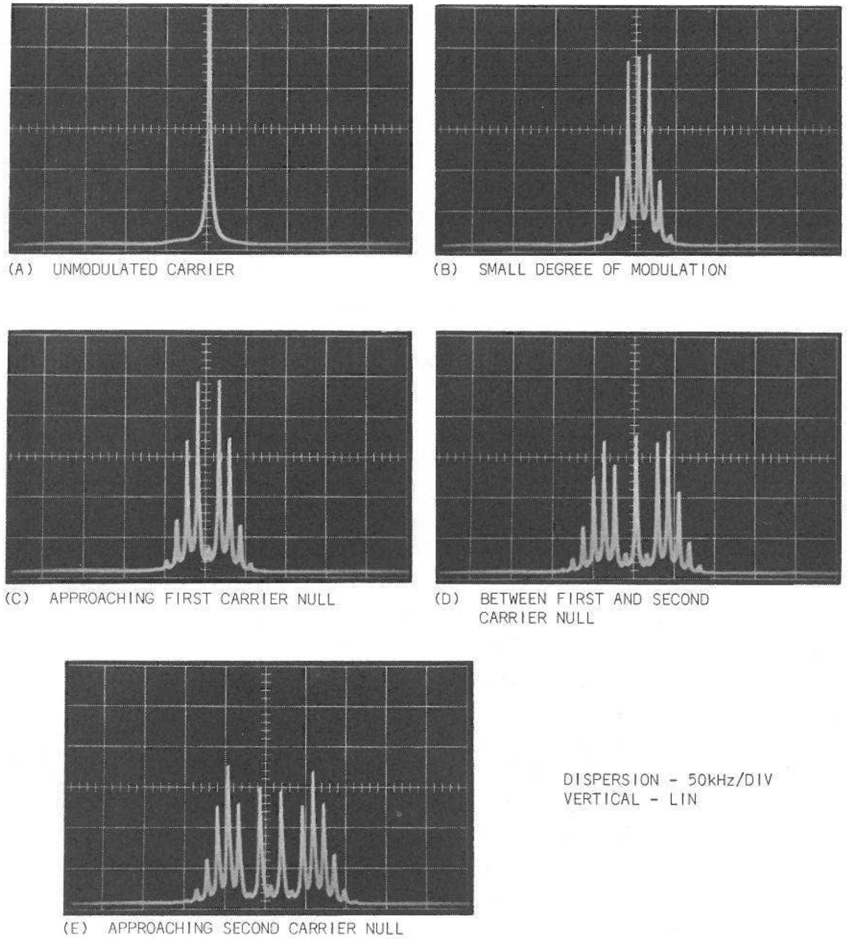


Fig. 8-9. Carrier-null method of deviation adjustment in FM using variable amplitude modulating signal.

With the modulating frequency set to 13,586 Hz by means of a counter, the modulating voltage is increased to obtain the second carrier null. This corresponds to a deviation of 75 kHz. Fig. 8-9A shows the unmodulated carrier corresponding to zero modulating voltage. As the modulating voltage is increased, the carrier amplitude decreases and sidebands appear as shown in Fig. 8-9B. As the modulating voltage is increased further, we reach the first carrier null corresponding to a modulation index of 2.4 and a deviation of 32.5 kHz. This is shown in Fig. 8-9C. As the modulating voltage is increased even further, the carrier amplitude increases again and more sidebands appear, as shown in Fig. 8-9D. Finally, in Fig. 8-9E with the modulating voltage increased yet again, the second null is reached corresponding to a frequency deviation of 75 kHz. As long as the final voltage setting is not exceeded, the transmitter will operate within the permitted limit of 75-kHz deviation.

EXAMPLE 3

measuring deviation linearity

Deviation linearity is a measure of the nonlinearity, existing in an FM transmitter or signal generator, between the carrier frequency deviation and the voltage amplitude of the modulating frequency causing the deviation. It is described graphically as the ratio of the modulating-frequency voltage divided by the modulation index to the modulation index, as shown in Fig. 8-10.

The measurement consists of measuring the voltage amplitude of the modulating frequency for successive carrier and sideband nulls for as many values of modulation index as desired. A

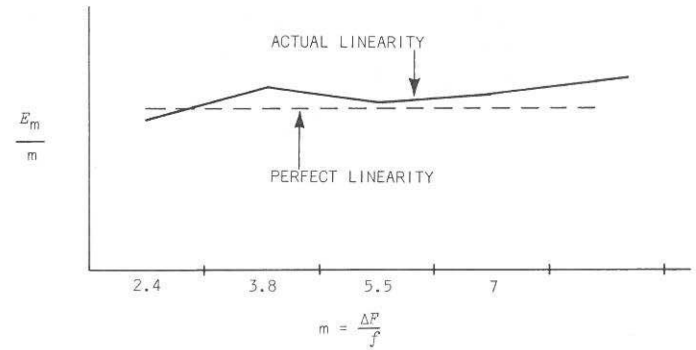


Fig. 8-10. Graphic display of deviation linearity.

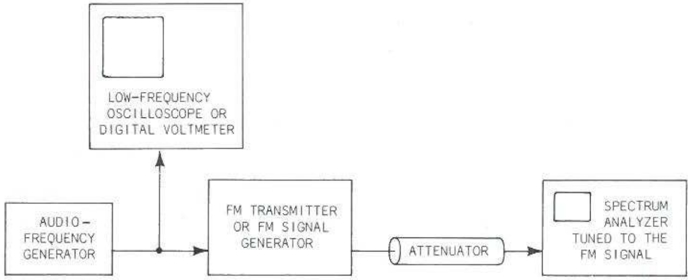


Fig. 8-11. FM-deviation linearity measurement.

spectrum analyzer capable of high resolution and low incidental FM is used to display the signal nulls, while the amplitude of the modulating frequency may be measured accurately with a low-frequency oscilloscope or digital AC voltmeter. See Fig. 8-11.

Ratios of modulating-frequency voltage divided by modulation index are plotted vertically and the values of modulation index are plotted horizontally to graphically display the degree of nonlinearity which may be present. Ideally, the curve should represent a horizontal straight line for the complete range of modulation index values.

Fig. 8-10 shows a graphical representation of the nonlinearity measured on a typical klystron high-frequency FM oscillator. The amount of nonlinearity is not affected by changing modulation frequencies.

using first-sideband null

Sometimes it is either inconvenient or impossible to obtain a carrier null. The method, then, is to null one of the sidebands, preferably the first sideband. The procedures are identical to those already described except that the modulation indices associated with the first sideband are used. Fig. 8-12 is a table of first-sideband nulls, and Fig. 8-13 is a spectrum-analyzer display showing the first first-sideband null at a modulation index of 3.83.

FIRST-SIDEBAND NULL	FIRST	SECOND	THIRD	FOURTH	FIFTH	SIXTH	SEVENTH	EIGHTH	NINTH
MODULATION INDEX $\frac{\Delta F}{f}$	3.83	7.02	10.17	13.32	16.47	19.62	22.76	25.90	29.05

Fig. 8-12. Table of first-sideband nulls.

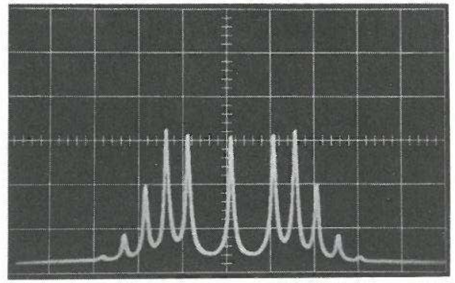


Fig. 8-13. FM-deviation measurement using null of first sideband.

ULTRAWIDEBAND FM

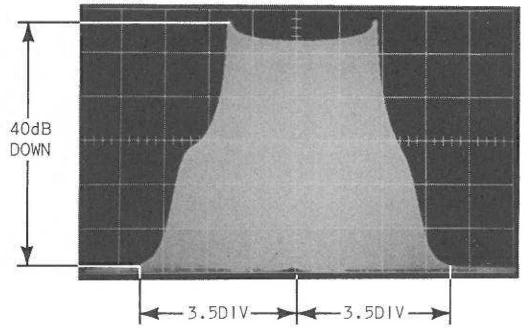
There are special problems in determining the frequency deviation of ultrawideband FM. These are:

1) Inability to resolve the separate frequency components of the signal because the modulation frequency is less than the narrowest spectrum-analyzer resolution bandwidth.

2) Even when the sidebands can be resolved, there is still difficulty in identifying the carrier among the many (sometimes hundreds) of displayed sidebands.

3) Even when the carrier is identified, it is almost impossible to count through more than about ten nulls without a large measure of uncertainty about the accuracy of the count.

The simplest method is to consider that the deviation is one half of the total occupied signal bandwidth as measured on the spectrum analyzer. This is based on the fact that the sideband energy falls off quite fast outside the frequency deviation. An approximate formula derived by Charest is $\frac{B}{\Delta F} = 2.0 + \frac{4}{\beta}$, where B is the 40-dB down bandwidth and β is the modulation index. Ignoring the $\frac{4}{\beta}$ term at modulation indices greater than 50 introduces less than a 5% error in the measurement. A more accurate simple formula for modulation indices less than 50 is $\frac{B}{\Delta F} = 2.5 + \frac{4}{\beta}$.



DISPERSION - 50kHz/DIV
VERTICAL - LOG
$\frac{1}{4}$ DIVISION FROM BOTTOM - 40dB DOWN

Fig. 8-14. Ultrawideband FM signal in frequency domain.

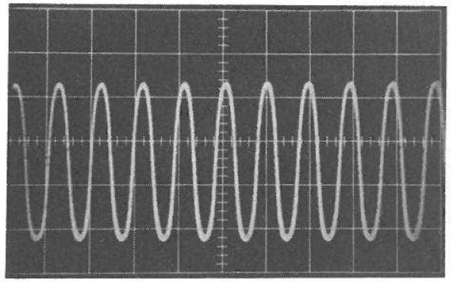
Fig. 8-14 shows a spectrum-analyzer display of an ultrawideband FM signal. Assuming that the deviation is one half the total 40-dB-down signal bandwidth, we obtain 7 div · 50 kHz/div = 350 kHz and the deviation is ΔF = 175 kHz. The modulating frequency is 100 Hz, which was determined from a different measurement as discussed later in this chapter. Here the modulation index is 1750, so that ignoring the $\frac{4}{\beta}$ term introduces negligible error. For a smaller modulation index, a correction factor computed from the $\frac{4}{\beta}$ term could be added to the original computation. More accurate calculations can be made by using more elaborate formulas. A detailed discussion on such measurements, including graphs, formulas, and sample calculations will be found in an article by C. N. Charest[3].

A note of caution on the signal bandwidth measurement: The accuracy of the 40-dB-down signal-width measurement depends on the resolution-bandwidth skirt selectivity. The wider the 40-dB-down resolution bandwidth, the greater the error. Unless the signal bandwidth is considerably greater than (e.g., 10 times) the resolution bandwidth, the effect of the resolution skirt selectivity should be considered.

[3]C. N. Charest, "Measuring Wide-Bandwidth FM Deviation," *EDN*, March 1, 1969.

DETERMINING MODULATION RATE FOR UNRESOLVED SIGNAL

Sometimes the modulation rate of a signal, either AM or FM, is less than the narrowest resolution bandwidth of the spectrum analyzer. This means that the modulating frequency cannot be obtained by the usual means of measuring the frequency difference between resolved adjacent signal components. In order to determine the modulating frequency, it is necessary to operate the spectrum analyzer as a time-domain superheterodyne radio receiver with a CRT indicator. This means that the sweeping oscillator is stopped by tuning to the zero-Hz/cm dispersion position. The modulation is now detected and displayed on the CRT. When the signal is AM, detection occurs at the peak of the resolution-amplifier resonance curve. When the signal is FM, detection occurs on the slope of the resolution-amplifier resonance curve as illustrated previously by Fig. 8-2. The modulation frequency is computed from the measured period of the displayed waveform. Fig. 8-15 shows the detected modulating waveform used in the ultrawideband FM signal of Fig. 8-14. Since the period of the waveform is 10 ms, the modulating frequency is 1/10 ms = 100 Hz.



DISPERSION - ZERO Hz/cm
VERTICAL - LIN
TIME BASE - 10ms/cm

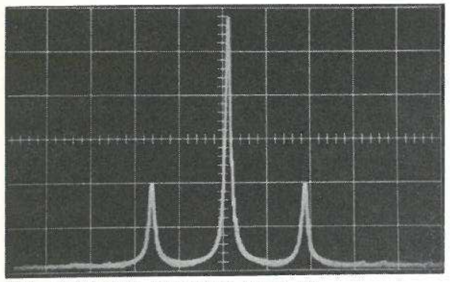
Fig. 8-15. Measuring modulating frequency at zero dispersion.

COMBINED AM AND FM

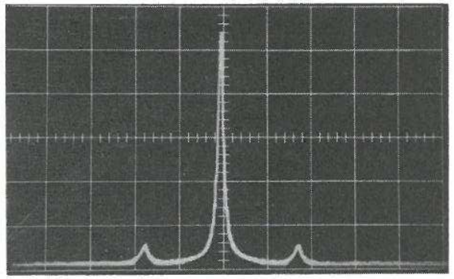
Combined AM and FM is usually an accidental, or incidental, occurrence. The desired modulation is usually AM, with the FM modulation an incidental by-product of an imperfect AM modulator. Combined AM and FM is characterized by two sidebands of unequal amplitude. This is because the AM sidebands are of the same phase while the FM sidebands are of opposite phase; for a detailed discussion see Chapter 4. Fig. 8-16 illustrates the measurement technique for combined AM and FM. Figs. 8-16A and 8-16B show the individual AM and FM spectra which, when generated simultaneously, result in the combined spectrum of Fig. 8-16C. Usually, only the combined spectrum, as shown in Fig. 8-16C, is available. From the unequal sideband amplitudes we conclude that the signal contains both AM and FM. Next, since the signal is supposed to be purely AM, we assume that the AM sidebands are larger than the FM sidebands. Except in very unusual circumstances this will always be the case. A further verification of the small size of the FM sidebands is the fact that the combined signal has only one significant sideband. We now compute the amplitudes of the individual AM and FM sidebands, using the fact that, in the combined spectrum, one sideband consists of the sum of an AM and FM sideband while the other sideband consists of the difference between an AM and FM sideband. From Fig. 8-16C, one sideband is about 2.3 cm high while the other is about 1.7 cm high. This leads to the conclusion that the AM sidebands are 2 cm high and the FM sidebands are 0.3 cm high. This is very close to the actual case as demonstrated in Figs. 8-16A and 8-16B. From the above we now compute:

$$\text{Percentage AM} = \frac{2 \cdot 2}{5.6}\,100 = 71.5\%$$

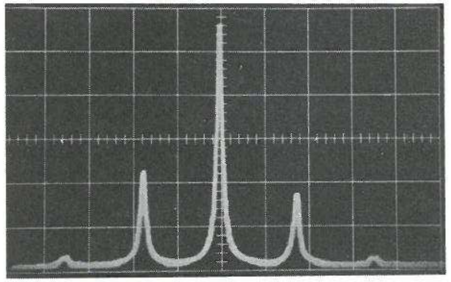
$$\text{FM Modulation Index} = \frac{2(0.3)}{5.6} = 0.107\,.$$



(A) SPECTRUM-ANALYZER DISPLAY OF AM



(B) SPECTRUM-ANALYZER DISPLAY OF NARROWBAND FM



(C) SPECTRUM-ANALYZER DISPLAY OF COMBINED AM AND FM

VERTICAL - LIN

Fig. 8-16. Combined AM and FM in the frequency domain.

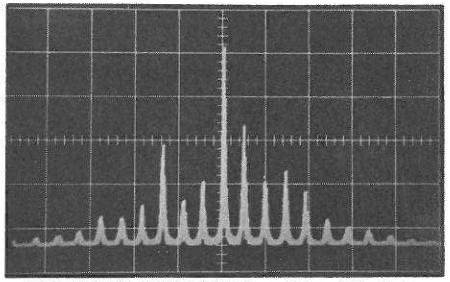


Fig. 8-17. Multitone FM in the frequency domain.

MULTITONE FM

Unlike AM, where there is no interaction between individual terms, the spectra of multitone FM are complex. These are usually not symmetrical, also additional sidebands at the sum and difference frequencies of the modulating signals may appear. The frequency-domain appearance of multitone FM is shown in Fig. 8-17. Observe that the spectrum is not symmetrical about the carrier component, which is at the center of the screen. Also note that there are additional sidebands besides those due to the 2-kHz and 3-kHz modulating frequencies. A detailed discussion of multitone FM will be found in Giacoletto's "Generalized Theory of Multitone Amplitude and Frequency Modulation."[4]

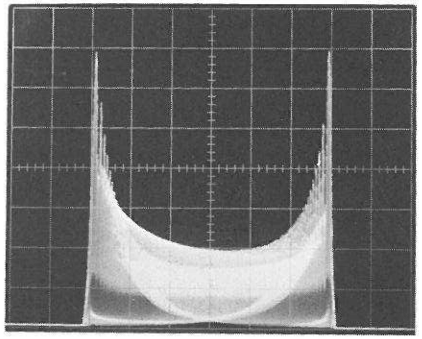


Fig. 8-18. Ultrawideband FM showing intensified portion in the center.

[4]Giacoletto, "Generalized Theory of Multitone Amplitude and Frequency Modulation," *Proc IRE*, July, 1947.

INTENSIFICATION EFFECTS

Frequently, ultrawideband FM spectra will have an intensified, or brighter, portion in the middle of the spectrum. Such a display is illustrated in Fig. 8-18. This intensification effect is due to the performance parameters of the spectrum analyzer, and is not indicative of anything about the signal. The effect is generated by the sweeping across the CRT screen of the resolution curve of the spectrum analyzer. As the FM signal frequency sweeps back and forth, so does the resolution curve of the spectrum analyzer. This back and forth sweeping movement causes two frequency interceptions to occur per FM peak-to-peak deviation. There are, however, two small frequency intervals at the ends of the deviation, where, because of the finite resolution bandwidth, there is only one interception. Consequently, the line density at the ends of the display is half as much as in the middle, causing a difference in brightness. Fig. 8-19 is a double exposure of two CW signals showing how the shape of the brightened portion of the spectrum is generated as the FM signal sweeps the resolution curve back and forth.

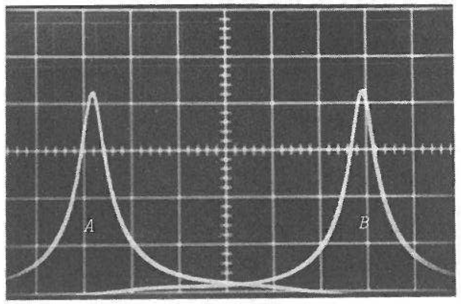


Fig. 8-19. As the resolution curve sweeps from *A* to *B* and back from *B* to *A*, the middle portion gets twice as many lines as the edges.

9

PULSED RF

The following basic relationships apply to the measurement of pulsed-RF signals.

1) Pulsed-RF measurements are usually based on the assumption of a *dense* rather than *discrete* type of spectrum. Accordingly, the discussion in this chapter applies only to dense spectra. Discrete spectra of pulsed signals will be considered under the heading of Waveform Analysis in Chapter 10.

2) In order to obtain a proper dense spectrum display, it is necessary that the resolution bandwidth B of the spectrum analyzer be greater than the pulse-train repetition rate. Mathematically:

$$B \geqslant \text{PRR}. \tag{9-1}$$

3) With condition (2) above established, the spectrum shape is traced by a series of vertical lines. These lines are not dispersion-dependent spectral lines. The vertical lines are sweep-speed-dependent repetition-rate lines. Each line represents one sample of the incoming signal. The number of lines on the CRT is equal to the number of pulses occurring during one spectrum-analyzer sweep.

4) In order to get sufficient definition of the spectrum shape, it is necessary to have a minimum of 5 sample, or rep-rate, lines per minor lobe and 10 lines for the major lobe. For a spectral display consisting of one major lobe and two minor lobes, this means twenty input pulses per spectrum-analyzer sweep. Hence, for proper definition of the spectrum shape:

$$\frac{20}{\text{PRR}} \leqslant 10\ (\text{time/div}) \tag{9-2}$$

5) The resolution of the fine details of the pulsed-RF spectrum depends on the resolution bandwidth used. The narrower the resolution bandwidth, the finer the details that can be observed. It has been found that the necessary fine details will be observed when:

$$t_0 B \leqslant 0.1, \tag{9-3}$$

where

t_0 = pulse width,

B = resolution bandwidth.

6) The spectrum-analyzer sensitivity is poorer for a pulsed-RF signal than for a continuous-wave signal having the same peak amplitude. The ratio in deflection height, on a linear voltage scale, between a pulsed-RF signal and a CW signal of equal peak amplitude is denoted by α.

$$\alpha = \frac{3}{2}\, t_0 B,$$

$$\alpha_{\mathrm{dB}} = 20 \log_{10} \frac{3}{2}\, t_0 B, \tag{9-4}$$

where

t_0 = pulse width,

B = 3-dB bandwidth

Since the pulsewidth-bandwidth product needs to be less than unity (see equation 9-3), it follows that α is invariably less than one, denoting a loss in sensitivity for pulsed signals.

7) Most pulsed-RF signals are of rectangular pulse shape. Examples of basic measurement techniques will, therefore, be based on rectangular pulses.

8) A detailed theoretical discussion of pulsed-RF will be found in Chapters 3 and 5. Fig. 9-1 shows the time-domain appearance of a pulsed-RF signal. From Fig. 9-1A, we observe that the period of the pulse train is 2(50) = 100 μs.

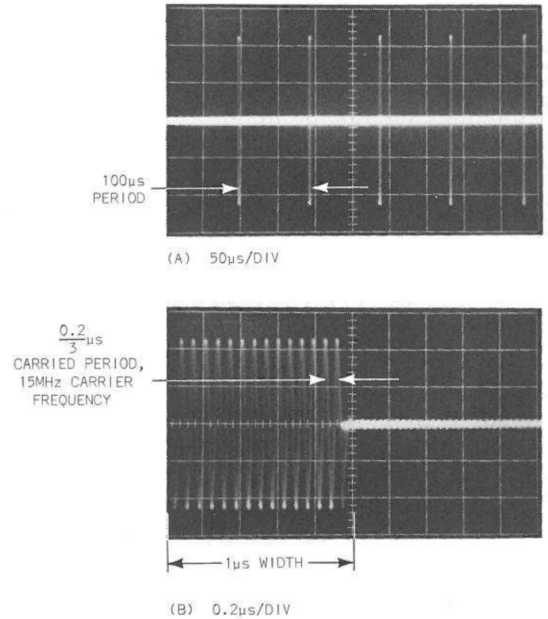


Fig. 9-1. Time-domain appearance of pulsed RF.

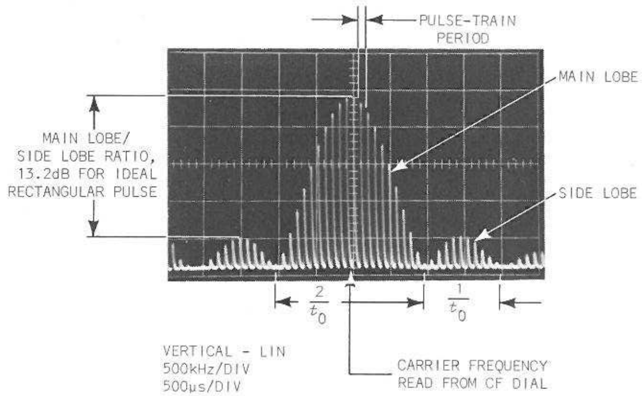


Fig. 9-2. Frequency-domain appearance of pulsed RF.

From Fig. 9-1B, we observe that the pulse width is 5(0.2) = 1 μs, the carrier frequency is $\frac{3}{0.2\ \mu s}$ = 15 MHz, and the pulse shape is rectangular. Fig. 9-2 shows the frequency-domain, spectrum-analyzer, appearance of the same pulsed-RF signal. This display contains the following information:

A) From the spectrum-analyzer center-frequency dial setting (not shown in Fig. 9-2), we ascertain that the carrier frequency is 15 MHz.

B) From the frequency width of the sidelobes, (2 div)(500 kHz/div) = 1 MHz, we compute that the pulse width is t_0 = 1/1 MHz = 1 μs.

C) From the time spacing between sample lines, we compute the pulse-train repetition rate or period. At 500 μs/div and 5 samples per division, the pulse-train period is 100 μs.

D) From the amplitude ratio of the mainlobe to first sidelobe, which is about 4.6/0.9, or 20 $\log_{10}$ 4.6/0.9 $\cong$ 14 dB, we conclude that the pulse shape is essentially rectangular. While 14 dB is greater than the theoretical 13.2 dB for a perfectly rectangular pulse, the difference is well within the measurement accuracy. Furthermore, a rectangular pulse gives the closest fit; for example, the ratio for a triangular pulse is over 26 dB.

Based on the above example, it is apparent that the same basic information can be obtained from both time-domain oscilloscope measurements and frequency-domain spectrum-analyzer measurements. The reason why spectrum analyzers predominate in this area is that, except in special cases, the oscilloscope is not able to display the signals in question. The limitations are frequency range (pulsed-RF signals in the gigahertz region are quite common) and sensitivity (many signals are in the picowatt region). Another difficulty with oscilloscope measurements is that frequently the desired information is the occupied frequency width or spectrum shape, rather than the time-domain pulse width or pulse shape. While it is theoretically possible to convert from one to the other by means of the Fourier mathematics, the task can be quite difficult.

radar performance

Most pulsed-RF measurements occur in radar systems. These involve both the determination of what a radar set is putting out, and the adjustment or calibration of the radar set so that it gives the required output. Spectrum analyzers are also frequently used in testing components such as pulsed magnetrons. The data of interest usually involves the following:

1) Carrier frequency (F),
2) Pulse width (t_0),
3) Pulse repetition rate (PRR), interpulse interval (T),
4) Pulse shape,
5) Occupied signal bandwidth,
6) Percentage missing pulses,
7) Carrier on/off ratio,
8) Presence of FM.

radar attenuation

Since radar sets usually put out much more power than the spectrum analyzer can accommodate, the signal connection is typically made through a directional coupler. Even then it may be necessary to add attenuation to the signal path to maintain spectrum-analyzer operation within the linear region. A typical test setup is shown in Fig. 9-3A. An alternate procedure, especially helpful in radar set tests as opposed to alignment, is to pick the transmitted signal off the air by a second antenna. Such an arrangement calls for a field transportable spectrum analyzer such as the portable Tektronix Type 491. Fig. 9-3B shows this alternate test arrangement.

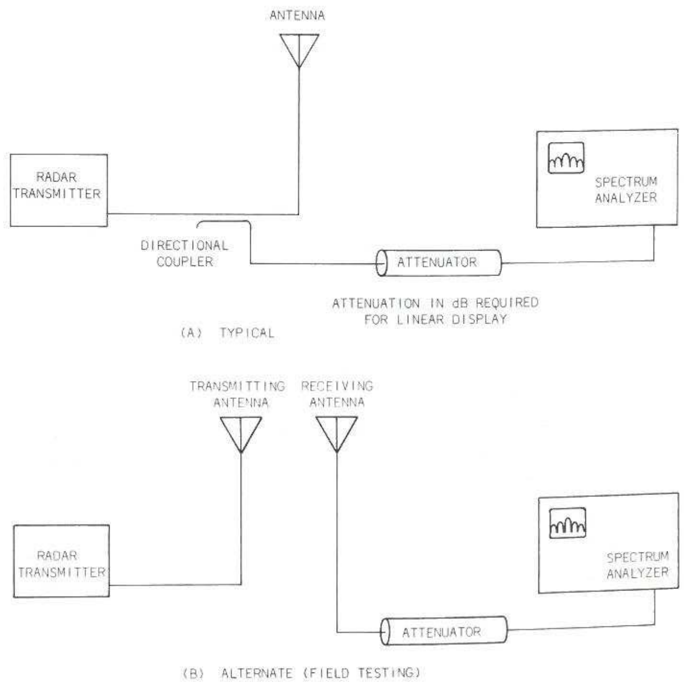


Fig. 9-3. Methods of measuring characteristics of a pulsed-RF spectrum generated by a radar transmitter.

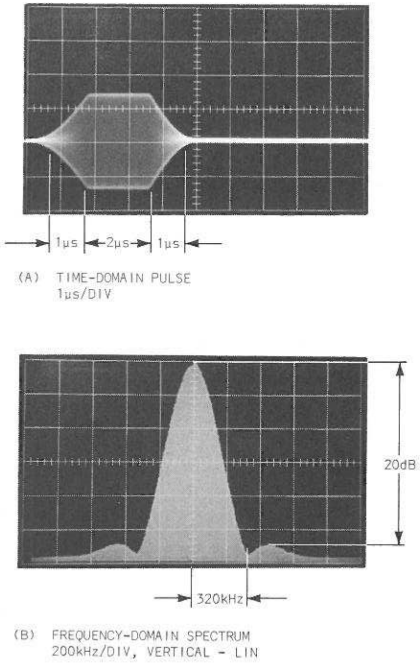


Fig. 9-4. Analyzing trapezoidal pulse.

EFFECT OF PULSE SHAPE

One of the more difficult characterizations of pulsed-RF in the frequency domain is that involving pulse shape. While it is theoretically possible to compute mathematically a time-domain shape from the frequency-domain spectrum, it is much easier to match the unknown spectrum against a previously established standard. The graphs of theoretical spectra as a function of pulse shape appearing at the end of Chapter 3 are quite useful for this purpose. While these graphs are based on theory, they are in close agreement with actual observations on a spectrum analyzer. This is illustrated by the following photographs.

trapezoid

Fig. 9-4A shows the time-domain shape of a trapezoidal pulse. It could also be described as a rectangular pulse having appreciable rise and fall time. Fig. 9-4B is the spectrum of this pulse. The ratio of mainlobe to first sidelobe in Fig. 9-4B is 6/0.6 or 20 log 10 = 20 dB. This is considerably more than the 13-14 dB expected for a rectangular pulse shape, but is short of the 26 dB expected for a triangular pulse. Thus, the pulse shape is somewhere between these two extremes. The choice of the precise shape, such as trapezoidal versus cosine-squared, cannot be made on the basis of Fig. 9-4B. For one thing, the spectrum analyzer does not indicate the spectrum *phase* characteristics. Fortunately, the user usually has an idea of what to expect. Once the user decides that he is dealing with a symmetrical trapezoid, the numbers are easy to compute. From Fig. 9-4B, the average pulse width (t_0) is the inverse of one-half the mainlobe width. Thus, $t_0 = \frac{1}{0.320} \cong 3.1\ \mu s$. From Fig. 9-4A, the average width is $2 + 2/2 = 3\ \mu s$, in good agreement with the computation.

triangle

Fig. 9-5 shows the time- and frequency-domain characteristics of a triangular pulse. Since the vertical display is logarithmic, the mainlobe to sidelobe ratio cannot be determined directly from the uncalibrated graticule. This number was determined to be 27 dB by inserting IF attenuation until the mainlobe deflection height was reduced to the former level of first-sidelobe deflection height. A sidelobe width of 300 kHz corresponds to the roughly 3-μs average pulse width. These numbers are in good agreement with theory.

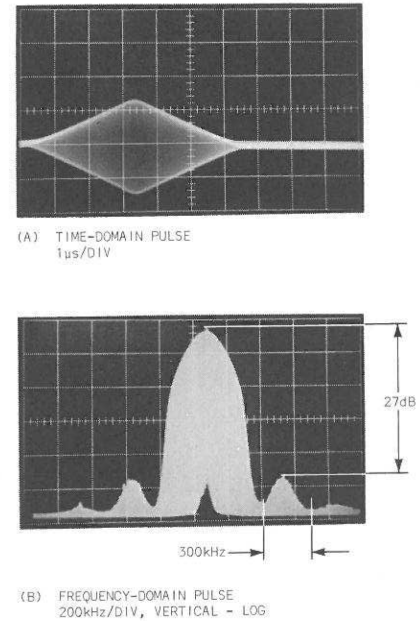


Fig. 9-5. Analyzing triangular pulse.

A somewhat different pulse-shape analysis problem is exemplified in Fig. 9-6. Fig. 9-6 is a double-exposure photograph showing the time- and frequency-domain characteristics of a sine-squared pulse. Such pulses are frequently used in the testing of television systems. One of the problems that arises is to ascertain how closely the pulse comes to being perfectly sine-squared. This is difficult to determine from time-domain testing. Small variations in pulse shape are, however, fairly obvious in the frequency domain, since these can have substantial effects on the spectrum. A graph of the theoretical spectrum of a sine-squared, or cosine-squared, pulse is shown in Table 3-2. Fig. 9-6 is in good agreement with the theoretically derived spectrum.

sine-squared

EFFECT OF FM

Pulsed-RF signals frequently contain a substantial amount of frequency modulation. The FM can be either intentional, such as in pulse compression radar, or, more usually, an unintentional byproduct of pulsing a magnetron or a klystron. In either case, the resulting spectra are different than those for non-FM'ing pulsed-RF signals. There are three major effects by which one can recognize the spectra of pulsed-RF of an FM signal from the spectra of pulsed-RF without FM:

A) Spectra of pulsed-RF signals without FM are always symmetrical, have distinct nulls, and the minor lobes are always smaller than the major lobe.

B) Spectra due to pulsed-RF of an FM'ing signal do not have distinct nulls; the sidelobes are larger than for non-FM'ing signals; for nonsymmetrical pulse shapes, such as a sawtooth, the spectrum is usually unsymmetrical.

The effect of FM on pulsed-RF was first described in *Radiation Laboratory Series*[1]. Recent experiments indicate that the amount of sidelobe lift-up is dependent on the product of pulse width and FM deviation. The greater this product the more the sidelobe lift-up.

[1] Montgomery, "Technique of Microwave Measurements," *Radiation Laboratory Series*, McGraw-Hill or Boston Technical Publishers, Vol. XI, Sec. 7.15.

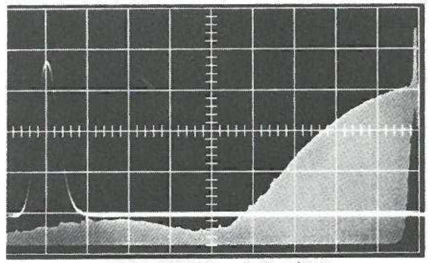


Fig. 9-6. Analysis of sine-squared pulse.

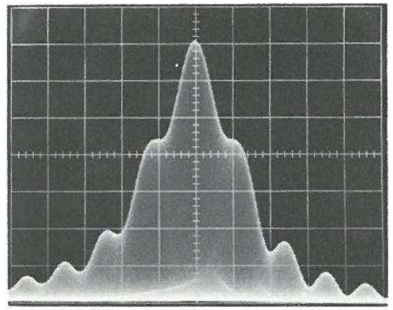


Fig. 9-7. Spectrum of rectangular pulsed-RF with FM.

pulsed-RF with FM

The following figures illustrate the frequency-domain appearance of pulsed-RF with FM. Fig. 9-7 is the spectrum of an FM'ing carrier having 6-MHz peak-to-peak FM deviation which is pulsed on in 5-μs intervals. Note that the inverse of the pulse width still determines the sidelobe width. Namely, $1/(5\ \mu s) = 0.2$ MHz. Similarly, carrier frequency and pulse repetition rate are determined the same way as for pulsed-RF without FM. However, the ratio of mainlobe-to-sidelobe size is much less, and the nulls between lobes no longer go down to zero. The pulse-width peak-to-peak FM-deviation product is:

$$t_0 \Delta F = 5\ \mu s \cdot 6\ \text{MHz} = 30.$$

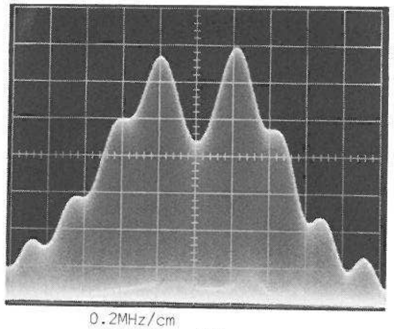


Fig. 9-8. Spectrum of rectangular pulsed-RF with FM.

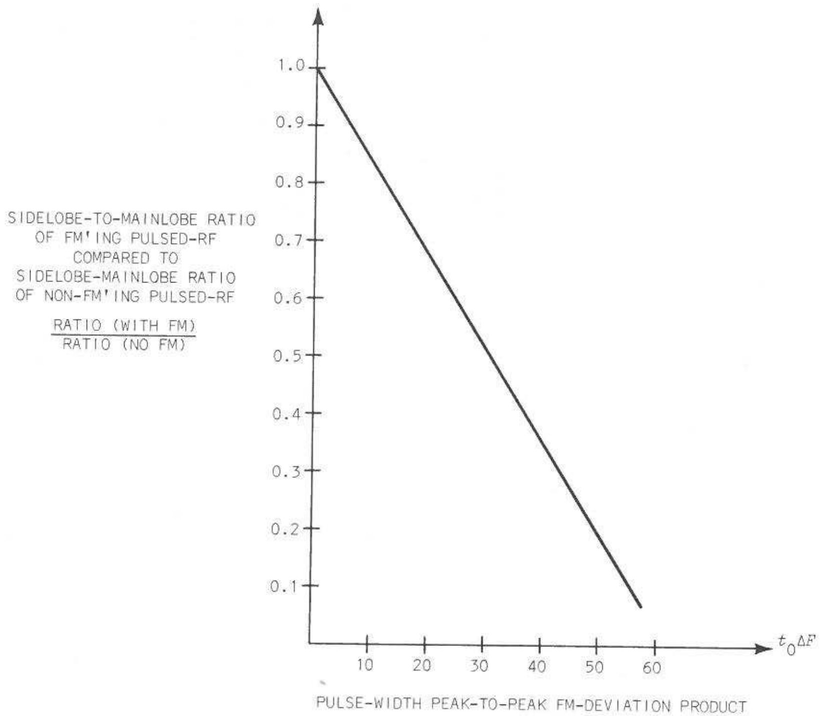


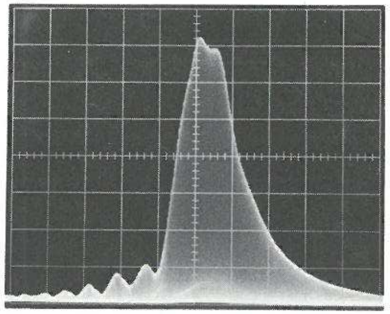
Fig. 9-9. Approximate experimental relationship between sidelobe-mainlobe ratio and peak-to-peak FM-deviation in FM'ing pulsed-RF.

Fig. 9-8 is a more extreme example of FM'ing pulsed-RF. Here the sidelobes are actually greater than the mainlobe. At a 5-μs pulse width and 10-MHz peak-to-peak FM deviation, the pulse-width peak-to-peak deviation product is 50.

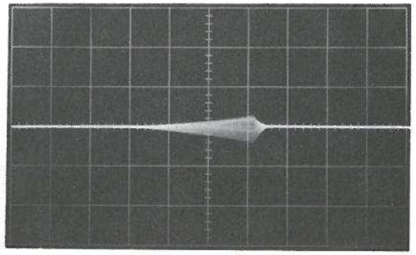
A plot of the data from Figs. 9-7 and 9-8, and data from recently published experiments[2], results in a close approximation to a straight-line relationship between sidelobe lift-up and pulse-width deviation product. Fig. 9-9 is such a graph where mainlobe-to-sidelobe ratio, normalized with respect to this ratio in the absence of FM, is plotted as a function of the pulse-width peak-to-peak FM-deviation product. The straight-line relationship is approximate, as it is based on insufficient data. This is, however, the best fit available at the present time. The purpose of the graph is to permit an estimate of the amount of FM present. The pulse width is determined in the normal manner from the inverse of the sidelobe width. The normalized mainlobe-to-sidelobe ratio is then determined by measuring the actual ratio and comparing it to the theoretical ratio of 4.6 (13.2 dB) for a rectangular pulse. The peak-to-peak FM deviation is then computed from the $t_0 \Delta F$ product read on Fig. 9-9. For example, the output of a pulsed-RF magnetron, using rectangular modulation pulses, yields a spectrum having a 3.5-to-one mainlobe-to-sidelobe ratio. The pulse width is 2 μs. What is the peak-to-peak FM deviation? The normalized mainlobe-to-sidelobe ratio is 3.5/4.6 = 0.76. From Fig. 9-9, this corresponds to a pulse-width total deviation product of about 14, which corresponds to a peak-to-peak FM deviation of 7 MHz for a 2-μs pulse width.

---

[2] Engelson & Breaker, "Spectrum Analysis of FM'ing Pulses," *Microwave Journal*, June, 1969.



(A) FREQUENCY DOMAIN
VERTICAL - LIN



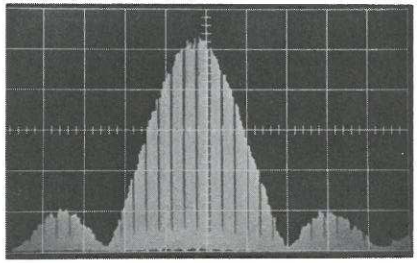
(B) TIME DOMAIN

Fig. 9-10. Spectrum of FM'ing asymmetrical pulse.

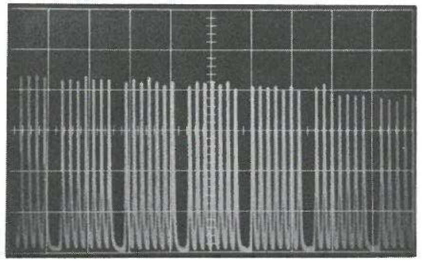
Fig. 9-10 shows the effect of FM on pulsed-RF having an asymmetrical pulse shape. The spectra of pulsed-RF, no matter what the pulse shape may be, is generally symmetrical. However, in the presence of FM, the spectra of pulsed-RF of asymmetric pulses are usually asymmetric. Fig. 9-10A shows the effect of FM on the spectrum of the asymmetrical triangular pulse of Fig. 9-10B. Note that not only have the minima been raised but the spectrum is highly asymmetrical.

PERCENTAGE OF MISSING PULSES DETERMINATION

One of the points of interest in pulsed-RF radar measurements is to determine the percentage of misfirings of the oscillator, usually a magnetron. The simplest way of doing this is by observation on a spectrum analyzer, where each of the vertical repetition-rate lines corresponds to an oscillator output pulse. The percentage of missing repetition-rate lines is, therefore, the same as the percentage of magnetron misfirings. Fig. 9-11 simulates the output of a misfiring oscillator. The deliberate misfirings shown in Fig. 9-11 are of a repetitive nature while in actual field situations these would be random, otherwise the two situations are the same. Fig. 9-11A shows the spectrum of a rectangular-pulse-shape pulsed-RF. It will be observed that periodically the signal disappears: these are misfirings. The percentage misfirings is determined by counting the repetition-rate lines. These lines are expanded across the CRT screen by going to a narrow, preferably zero-hertz, dispersion position and a sweep time which permits counting individual lines. This is shown in Fig. 9-11B. Here it will be observed that one out of every eight lines is missing. The percentage misfirings, therefore, is $1/8 \cdot 100 = 12.5\%$. In a real situation, the misfirings would be random. Hence, several photos would have to be taken in order to get a statistically significant number of misfirings.



(A) OVERALL SPECTRUM



(B) EXPANDED REPETITION-RATE LINES

Fig. 9-11. Measuring percentage of oscillator misfirings.

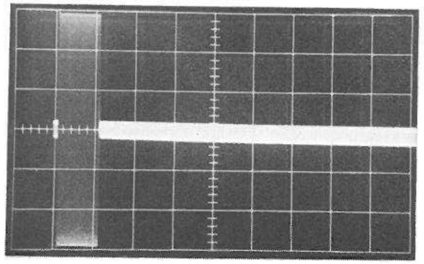


Fig. 9-12. Pulsed-RF, poor on/off ratio in time domain, 2 μs/div.

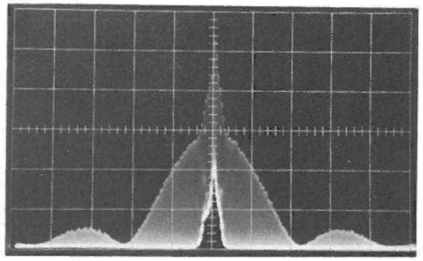


Fig. 9-13. Pulsed-RF, poor on/off ratio in frequency domain, 200 kHz/div, vertical – LIN.

MEASURING MODULATOR ON/OFF RATIO

Frequently the modulator or oscillator generating the pulsed-RF waveform will not turn off completely and there continues to be a small amount of output during the interpulse interval. This leads to the need for determining the on/off ratio of the equipment. Fig. 9-12 shows the time-domain appearance of a pulsed signal with a poor on-to-off ratio. The *on* amplitude is close to 6 divisions while the *off* amplitude gives a deflection of 0.4 divisions; hence, the on/off ratio is 6/0.4 = 15, or 20 log 15 = 23.5 dB. To determine the on/off ratio from frequency-domain measurements, we proceed as follows:

1) Fig. 9-13 is a frequency-domain display of the spectrum corresponding to Fig. 9-12. The spectrum consists of the superposition of two parts; one is basically a typical pulsed-RF spectrum while the other part is an ordinary CW spectrum[3]. We observe that in Fig. 9-13 the CW part of the spectrum consists of a two-division deflection. The CW response shows up twice: once as a standard resolution curve (the hole in the middle) and again as an addition on top of the pulsed-RF spectrum. Subtracting the effect of the CW response from the overall display, we find that the mainlobe of the pulsed-RF spectrum has a three-division deflection.

2) Having determined the deflection amplitudes of the CW and pulsed-RF parts of the spectrum, their ratio can be determined; in our example this is 3/2 or 3.5 dB. Note that what is significant is the *ratio* of the deflection amplitudes rather than the actual deflection amplitudes. For large on/off ratios, it may be necessary to make this measurement in the LOG mode. Under these conditions the ratio, or difference in dB, is determined either from a calibrated graticule or by using the calibrated IF attenuator.

3) The computation of part two is only part of the answer. This is because the frequency-domain display does not show the true relative amplitude difference between the CW and pulsed part of the signal. For pulsed-RF, there is a loss in sensitivity relative to CW signals. This loss in sensitivity is given by equation (9-4), it is:

$$\alpha = \frac{3}{2} t_0 B,$$

where

t_0 = pulse width,

B = 3-dB bandwidth.

It is therefore necessary to determine the pulse width and the 3-dB bandwidth before proceeding further.

[3] For poor on/off ratios we cannot use the idea of simple superposition. This is because the part of the signal causing the CW spectrum makes the actual pulse height smaller. For on/off ratios greater than about 20 dB, the error due to this effect is negligibly small.

4) From Fig. 9-13, the sidelobe width is 2.5 · 200 = 500 kHz. The pulse width is computed:

$$t_0 = \frac{1}{0.5} = 2\ \mu s.$$

5) If the spectrum analyzer in use has calibrated resolution bandwidths, this number is read directly from the front panel, otherwise it has to be measured as follows: Without changing the resolution bandwidth setting, apply a CW signal to the spectrum analyzer. The center frequency or dispersion setting has no effect on this measurement so any frequency CW signal may be used. If a CW signal is not available, the CW feedthrough part of the original signal may be used, though this makes a more difficult measurement. Fig. 9-14 illustrates how the 3-dB bandwidth is measured. After obtaining a convenient deflection height, the signal amplitude is increased by 3 dB and the display width is measured at the deflection height of the original reference point. To perform this measurement, it is not necessary to have a calibrated signal generator. The 3-dB amplitude change can be made by using the calibrated attenuator which is found in every modern spectrum analyzer. From Fig. 9-14, the display width at the 3-dB-down point is 0.65 divisions. At 50 kHz/div, the 3-dB bandwidth is (50)(0.65) = 32.5 kHz.

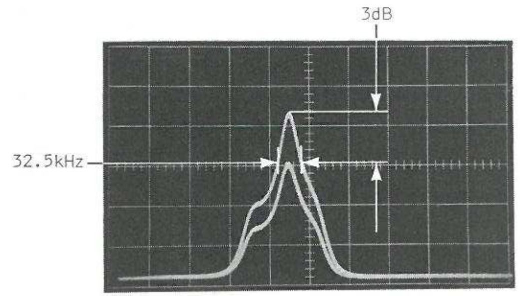


Fig. 9-14. Measuring the 3-dB-down resolution-amplifier bandwidth, 50 kHz/div, vertical – LIN.

6) We can now compute the loss in pulse sensitivity which is:

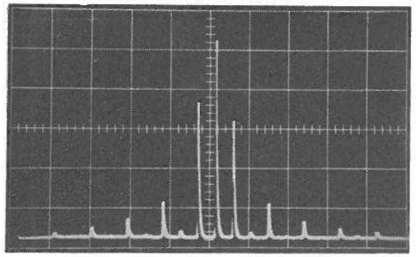
$$\alpha_{dB} = 20 \log (1.5)(32.5 \cdot 10^3)(2 \cdot 10^{-6})$$

$$= 20 \log 0.0975 = -20.2\ \text{dB},$$

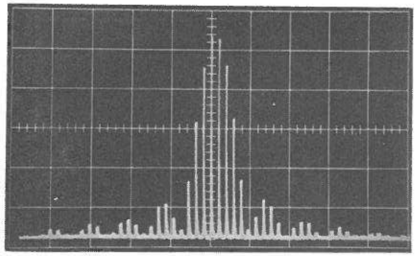
where the minus sign denotes a loss.

7) The total on/off ratio is the sum of the two computations; namely, on/off ratio = 20.2 + 3.5 = 23.7 dB. This is in good agreement with the 23.5 dB computed from time-domain data.

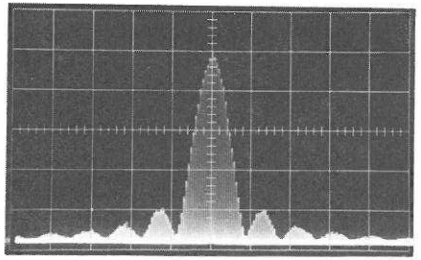
Note that when we added the two numbers, we added two losses to find the total loss in sensitivity.



(A) 0.5ms/cm



(B) 1ms/cm



(C) 3ms/cm

Fig. 9-15. Effect of sweeptime on spectrum definition. Rectangular pulsed-RF, 5-kHz rate.

EFFECT OF CONTROL SETTINGS

REPETITION RATE

Before any measurements can be undertaken, it is necessary to have enough repetition-rate samples to define the overall shape of the spectrum. Fig. 9-15 illustrates this. Fig. 9-15A has insufficient rep-rate lines for shape definition. In Fig. 9-15B, the overall shape of the spectrum is just becoming apparent, while the shape in Fig. 9-15C is very clearly defined. A count of repetition-rate lines in Fig. 9-15B will show 5 sample lines per minor lobe and 10 lines for the major lobe. This is usually considered the demarcation line between a defined and undefined spectrum shape. Since the number of lines on the screen is equal to the number of pulses intercepted during one sweep, it is necessary to sweep slower (more time per cm) if there are insufficient lines to define the spectrum shape.

DENSE VERSUS LINE SPECTRUM

Pulsed-RF spectrum analyzer measurements are usually based on a dense-spectrum rather than line-spectrum interpretation. To achieve a dense-spectrum type of display, it is necessary that the spectrum-analyzer resolution setting be greater than the pulse repetition rate. This matter is discussed in considerable detail in Chapter 5. Fig. 9-16 illustrates what happens to the appearance of the spectrum as the relationship between resolution bandwidth and pulse repetition rate changes. Fig. 9-16A shows a standard pulsed-RF spectrum. To obtain this display it was necessary that the resolution bandwidth be greater than the pulse repetition frequency. As the pulse repetition rate is increased, we reach a point where it becomes equal to the resolution bandwidth. This is shown in Fig. 9-16B. This is the transitional spectrum between the dense display of Fig. 9-16A and the line, or CW, display of Fig. 9-16C. In Fig. 9-16C, the pulse repetition rate has been increased to several times the resolution bandwidth.

The following evolution in the spectrum is observed as the pulse repetition rate is increased. At first, more sample lines appear on the screen but the spectral shape remains unchanged. The spacing of the lines depends only on the time-per-division setting and is independent of the dispersion setting. This is the normally desired mode of operation. As the pulse repetition rate becomes equal to, and then exceeds, the resolution bandwidth setting, the spectrum amplitude increases and the lines no longer go down to the baseline. This is the transitional stage. Eventually, new lines going down to the baseline appear. These look like and behave like ordinary CW signal displays. This shape is easily recognized by the fact that spacing between lines is independent of the sweep time and is determined solely by the dispersion setting. In Fig. 9-16, the spectrum-analyzer gain was progressively reduced, in going from (A) to (C), so as to maintain the display on the screen.

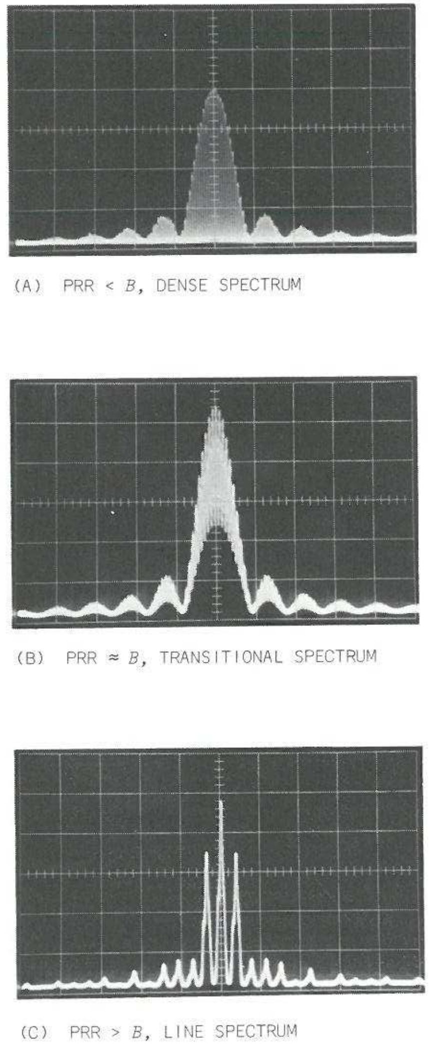
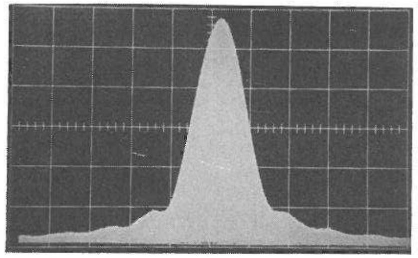
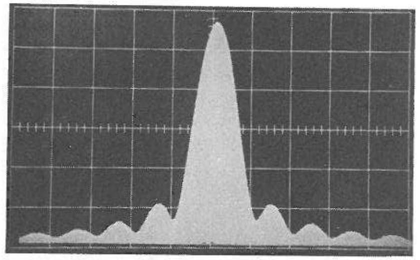


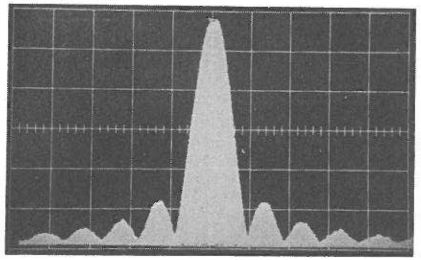
Fig. 9-16. Effect of the pulse-repetition-rate resolution-bandwidth relationship on the type of spectrum.



(A) $t_0B \geq 0.5$



(B) $t_0B \approx 0.15$



(C) $t_0B < 0.1$

Fig. 9-17. Effect of pulsewidth-bandwidth product on spectral display.

FINE DETAIL

In order to display the fine detail of a pulsed-RF spectrum, it is necessary that the pulsewidth-resolution-bandwidth product be less than one tenth. Mathematically: $t_0B \leqslant 0.1$. The effect of not meeting this requirement is illustrated in Fig. 9-17. In Fig. 9-17A, the pulsewidth-bandwidth product is considerably greater than one tenth. Note that the sidelobes are almost completely obscured and there are no nulls. In Fig. 9-17B, the pulsewidth-bandwidth product is slightly greater than one tenth. Here the sidelobes are clearly outlined and the position of the nulls is definite. This permits the unambiguous determination of pulse width. The nulls are, however, not very deep. This makes it difficult to ascertain the degree of incidental FM present. In Fig. 9-17C, where the pulsewidth-bandwidth product is slightly below one tenth, the nulls are sharp and clear. This not only indicates that the spectrum is properly resolved but also that no incidental FM is present.

EFFECT OF SENSITIVITY AND DYNAMIC RANGE

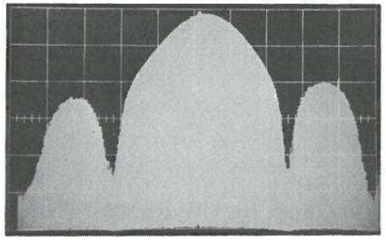
As previously indicated, the sensitivity for pulsed-RF signals is less than for CW signals. The ratio of display height on a linear scale, or ratio of sensitivities, between pulsed signals and CW signals of equal peak amplitude is given by the formula:

$$\alpha = \frac{3}{2} t_0B.$$

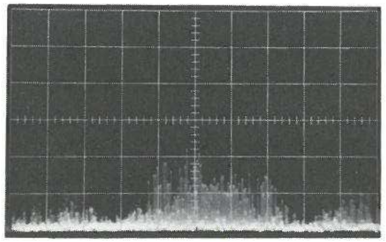
Since, for appropriate display of fine spectrum detail, it is necessary that $t_0B \leqslant 0.1$, it follows that for a properly displayed pulsed-RF signal $\alpha \leqslant 0.15$. An alpha of 0.15 represents a 16.5-dB loss in sensitivity. This is the minimum possible loss in pulsed-RF versus CW sensitivity that will still permit a correct display of the spectrum. This decrease in sensitivity would not cause much difficulty except that it also degrades the instrument dynamic range. The loss in dynamic range occurs because the peak input power level for linear operation is little affected by pulse width. This fundamental relationship limits how narrow of a pulse may produce an analyzable spectrum. The narrowest pulse width is determined by the widest resolution bandwidth available. For example, for the Tektronix Type 1L20 Spectrum

Analyzer, the widest resolution bandwidth is 100 kHz. All pulse widths less than 1 μs ($t_0 B < 0.1$) will produce sensitivity losses greater than the theoretical minimum of 16.5 dB. Eventually, the loss in sensitivity, and accompanying loss in dynamic range, will preclude the display of a measureable spectrum.

This measurement limitation is illustrated in Fig. 9-18. Fig. 9-18A shows the spectrum of a 0.2-μs pulse using a 100-kHz resolution bandwidth. In spite of the $\alpha = 1.5 \cdot 0.2 \cdot 10^{-6} \cdot 10^5 \rightarrow 30.5$-dB loss in sensitivity and dynamic range, the spectrum can still be displayed over the full 40-dB LOG dynamic range of the spectrum analyzer. Fig. 9-18B shows the spectrum of an 80-ns pulse. Here only two centimeters of vertical drive could be achieved, indicating that pulse widths narrower than about 0.1 μs cause too much loss in dynamic range.

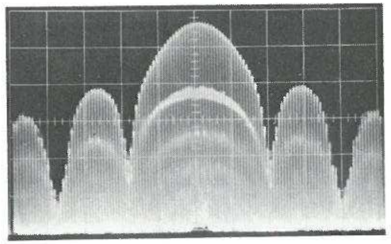


(A) 2MHz/cm
100kHz BANDWIDTH
0.2μs PULSE WIDTH
VERTICAL - LOG

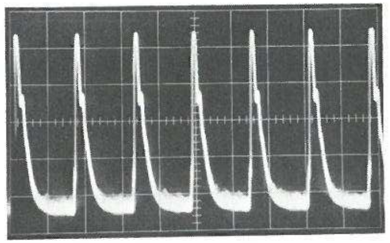


(B) 5MHz/cm
100kHz BANDWIDTH
80ns PULSE WIDTH
VERTICAL - LOG

Fig. 9-18. Illustrating the loss in dynamic range for narrow pulses.



(A) PULSED-RF
VERTICAL - LOG



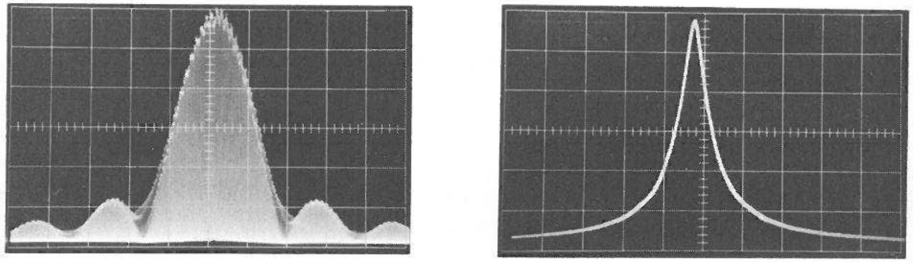
(B) PULSED-RF
ZERO Hz/cm
VERTICAL - LOG

Fig. 9-19. Pulsed-RF illustrating bands of intensification across spectrum.

DISPLAY INTENSIFICATION EFFECTS

Sometimes the spectral display may have extraneous features that are not readily accounted for. Frequently, the peculiarity is in the form of a brightening or intensification of a portion of the spectrum. These intensification effects are almost always due to the operating parameters of the spectrum analyzer. The phenomenon gives no information about the signal, and so should be ignored. Two such effects will now be illustrated and causes explained.

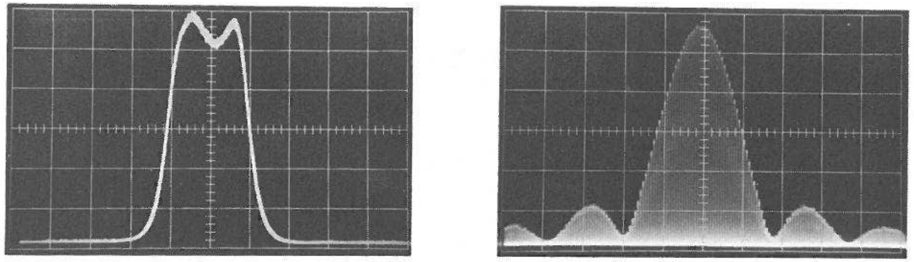
Fig. 9-19A is the spectrum of a pulsed-RF signal. Note the several bands of intensification, the most prominent being two centimeters from the top. This intensification is caused by the transient response of the variable resolution amplifier.



(A) PULSED-RF
50kHz/cm
VERTICAL - LIN

(B) RESOLUTION AMPLIFIER RESPONSE
FOR (A)
5kHz/cm
VERTICAL - LIN

Fig. 9-20. Effect of gradual-resolution-filters skirt selectivity on pulsed-RF spectrum.



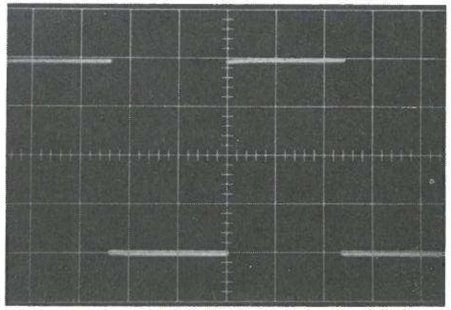
(A) VERTICAL - LIN

(B) VERTICAL - LIN

Fig. 9-21. Sharp skirt selectivity leads to well defined nulls.

Fig. 9-19B is an expanded version of Fig. 9-19A. We see that each of the repetition-rate sample lines, which are due to the variable-resolution-amplifier transient response as explained in Chapter 5, has an overshoot two centimeters from the top of the display. These overshoots, which cannot be resolved individually in Fig. 9-19A, give the effect of a more intense trace.

A different type of intensification effect is illustrated by Fig. 9-20A and appears to be two spectra. One, quite bright, appears to be a standard $(\sin x)/x$, representative of a rectangular pulsed RF. The other, somewhat less intense spectrum, has no sidelobes or minima. This effect is due to gradual rather than abrupt skirts on the variable resolution filter as illustrated in Fig. 9-20B. From Fig. 9-20A, we calculate that the pulse width is roughly 11 μs. In order to have adequate fine-detail definition, it is necessary to meet the requirement that $t_0 B < 0.1$. With $t_0 = 11\ \mu$s, the bandwidth has to be less than 9.1 kHz. The resolution curve is quite adequate at the 3-dB or even 6-dB down point. But, at the amplitude level of the spectrum nulls, the resolution curve is about 20 kHz wide. This leads to the combination spectrum of Fig. 9-20A. When the resolution curve shape is more rectangular, Fig. 9-21A, the nulls are clearly defined, as shown in Fig. 9-21B. Once the cause of the double spectrum in Fig. 9-20A is understood, the filled-in sidelobes can be ignored and all necessary data obtained.



HORIZONTAL - 20μs/cm
VERTICAL - 0.2V/cm

Fig. 10-1. Time-domain appearance of squarewave.

10

MISCELLANEOUS APPLICATIONS

WAVEFORM ANALYSIS

Fourier-series theory, as discussed in Chapter 3, provides a mathematical relationship between the time-domain and frequency-domain characteristics of various waveforms. This means that one can compute the time-domain characteristics of a wave train from frequency-domain measurements and vice versa. The measurement and computational technique is illustrated by the following examples.

squarewave

Fig. 10-1 shows the time-domain appearance of a squarewave. The period of the squarewave is 5cm · 20μs/cm = 100 μs and the peak-to-peak amplitude is 4cm · 0.2V/cm = 0.8 V. According to Fourier theory (see Table 3-2), the squarewave is composed of sinewaves whose amplitudes are given by

$$C_n = \frac{2At_0}{T} \frac{\sin \frac{n\pi t_0}{T}}{\frac{n\pi t_0}{T}}, \tag{10-1}$$

where C_n is the zero-to-peak ($\sqrt{2}$ RMS) amplitude of the nth harmonic of the sinewaves whose superposition makes up the squarewave. The fundamental sinewave frequency (n = 1) has the same period as the squarewave. Hence,

$$f_0 = \frac{1}{T} = \frac{1}{100\ \mu s} = 10\ \text{kHz}.$$

For a symmetrical squarewave, the ratio of pulse width to period is t_0/T = 1/2. Hence, the amplitude of the fundamental is:

$$C_1 = \frac{2A}{2} \frac{\sin \frac{\pi}{2}}{\frac{\pi}{2}}.$$

For our example, we get

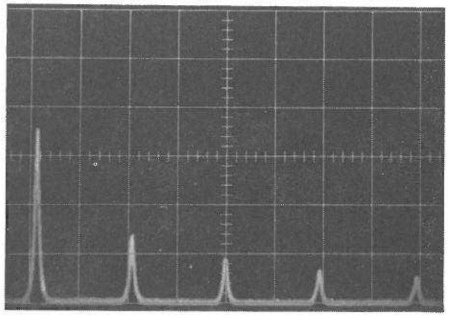
$$C_1 = \frac{(0.8)(1)(2)}{\pi} = 0.51 \text{ V zero to peak,}$$

$$C_2 = \frac{(0.8)(\sin \pi)(2)}{2\pi} = 0, \text{ since } \sin \pi = 0,$$

$$C_3 = \frac{(0.8)(1)(2)}{3\pi} = 0.17 \text{ V zero to peak.}$$

Continuing in like manner, we observe that all even harmonics are zero, while the odd harmonics decrease in amplitude in proportion to the harmonic number.

The qualitative information on the shape of the time-domain waveform can be determined by observing the spectrum on a spectrum analyzer having an appropriate frequency range. Thus, the absence of even harmonics and the fact that the third harmonic is one third as large as the fundamental, the fifth harmonic being one fifth of the fundamental, etc., indicates that we are dealing with a squarewave. The quantitative information dealing with frequencies and amplitudes requires a fully calibrated spectrum analyzer, such as the Tektronix Type 1L5. Fig. 10-2 shows the frequency-domain appearance of the squarewave as resolved by a Type 1L5 Spectrum Analyzer. The frequency spacing of the fundamental and harmonics shows that the repetition rate of the original waveform is about 10 kHz.



HORIZONTAL - 10kHz/cm
CENTER - 50kHz
VERTICAL - LIN, 0.1V/cm RMS

Fig. 10-2. Frequency-domain appearance of squarewave.

Missing even harmonics and the relative amplitude of the remaining harmonics indicate a squarewave. The amplitude of the fundamental comprising the squarewave is 3.6cm · 0.1V/cm = 0.36 V RMS, or $0.36\sqrt{2}$ = 0.51 V zero to peak. Conversely, the time-domain peak-to-peak value can be computed from:

$$C_1 = \frac{2A}{\pi} \text{ zero to peak}$$

$$= \frac{\sqrt{2}A}{\pi} \text{ RMS,}$$

or

$$A = \frac{C_{1(\text{RMS})}\,\pi}{\sqrt{2}} \text{ V peak to peak} \qquad (10\text{-}2)$$

Substituting for C_1 from Fig. 10-2 we get

$$A = \frac{0.36\pi}{\sqrt{2}} = 0.8 \text{ V peak to peak,}$$

as measured from Fig. 10-1.

In actual practice, one would normally not use a spectrum analyzer to evaluate characteristics of a squarewave. The job is more easily accomplished in the time-domain with an oscilloscope. There are, however, some rare cases where the spectrum analyzer is better. Such a situation might arise where the squarewave amplitude is insufficient to show up on the oscilloscope but can be observed on the more sensitive spectrum analyzer. The major spectrum-analyzer application in this area is as a tool illustrating the physical meaning of the theoretical Fourier series analysis. The spectrum analyzer can also be used as a device for the direct determination of the Fourier series of waveforms in lieu of a complicated computation.

A highly useful spectrum-analyzer application in the area of waveform analysis is that of symmetry adjustment. Fig. 10-3 shows an oscilloscope display of a slightly asymmetrical squarewave. The asymmetry is hardly noticeable. Fig. 10-4 shows the same squarewave in the frequency domain. Here, the presence of even harmonics is a clear indication of asymmetry. The signal source could be tuned for best symmetry by adjusting for minimum even-harmonic generation.

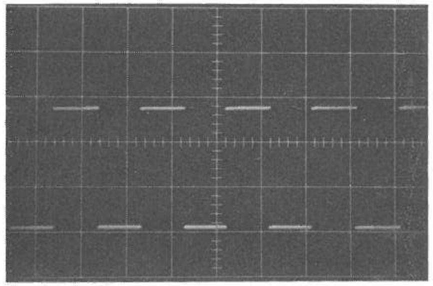


Fig. 10-3. Slightly asymmetric squarewave in time domain.

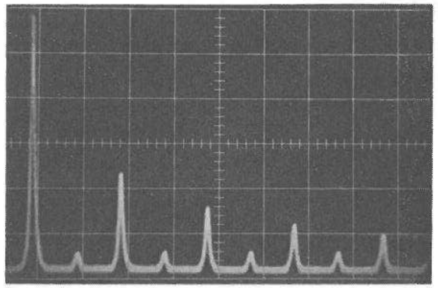


Fig. 10-4. Slightly asymmetric squarewave in frequency domain. Note even harmonics.

When making computations in waveform analysis, it is important to keep in mind the calibration factors of the instruments and the units of the formulas. Thus, oscilloscopes usually yield peak-to-peak values, spectrum analyzers are usually calibrated in RMS, and the Fourier series formulas are based on zero-to-peak levels. Neglecting to convert to a common set of units will result in an error of either $2\sqrt{2}$ or $\sqrt{2}$.

RANDOM NOISE MEASUREMENT

The subject of random noise is quite complex. A complete discussion of this subject requires probability mathematics – a subject which is beyond the scope of this volume. Those interested will find many references in this area[1]. The frequency-domain description of noise is expressed by a power spectrum where the basic element is *spectral density* in units of power per unit bandwidth. Besides the need to express noise in units of power per bandwidth, the absolute level indicated by a measuring device is also affected by the detector time constant.

The effect on the output of various detector time constants and a summary of useful formulas will be found in the cited references[2].

However, the above complications do not arise if one is willing to settle for a partial, relative-distribution, expression of the signal. A relative-amplitude-distribution characterization is all that is needed for many applications. This can be easily obtained with any spectrum analyzer.

[1] See, for example, Davenport & Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, 1958; also, Blackman & Tukey, *The Measurement of Power Spectra*, Dover Publications, 1958.

[2] Peterson, "Response of Peak Volt-Meters to Random Noise," *GR Experimenter*, Vol. 31, No. 7, December, 1956; also, "Useful Formulas, Tables, and Curves for Random Noise," GR *Instrument Notes*, IN-103, 1967.

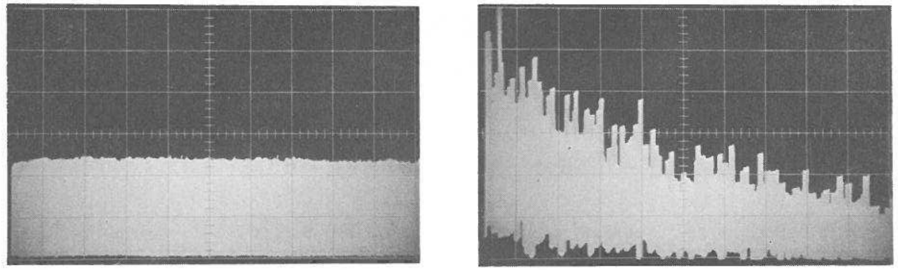
Figs. 10-5 and 10-6 illustrate such a measurement. The illustrated problem was to adjust the bias on a zener noise source for the flattest output frequency distribution. The two oscilloscope presentations, Figs. 10-5A and 10-6A, show relative output amplitude, but the frequency distribution cannot be ascertained. The spectrum-analyzer displays, Figs. 10-5B and 10-6B, clearly show that the bias adjusted for Fig. 10-6 results in the flatter output frequency distribution.

DISTORTION MEASUREMENT

A frequent spectrum-analyzer application is to determine the degree of distortion in what is supposed to be a sinusoidal waveform. The method consists of measuring the relative amplitudes between the fundamental and the various harmonics and computing the percentage of harmonic content. When there is more than one significant harmonic, the percentages are usually combined by the RMS-sum method – that is, by taking the square root of the sum of the squares. Fig. 10-7 illustrates a typical percentage distortion measurement. Fig. 10-7A is the oscilloscope presentation of a signal source output; the waveform is obviously not a perfect sinewave. The degree of distortion is, however, very difficult to ascertain from this presentation. Fig. 10-7B is a spectrum-analyzer display for the same waveform. From the spectrum-analyzer display, we observe that the several harmonics are 25 dB, 28 dB, 26 dB, and 36 dB below the fundamental. Converting from dB to voltage ratios, we get: 1/17.8, 1/25.2, 1/20, and 1/63.3 as the amplitudes of the harmonics with respect to the fundamental.

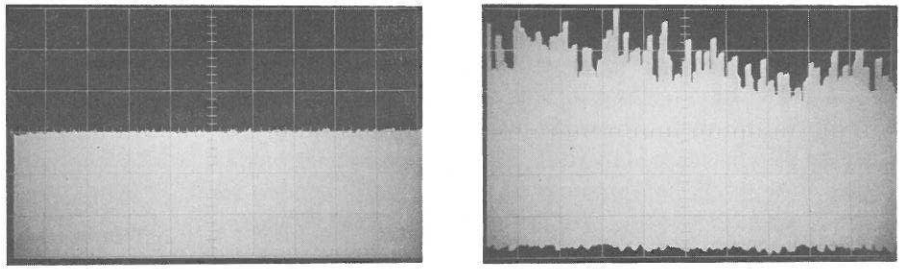
The percentage harmonic content is obtained by multiplying each ratio by 100. Thus, the respective harmonic percentages are: 5.6%, 4.0%, 5.0%, 1.6%. The total harmonic distortion (HD), computed by the RMS-sum method, is:

$$HD = \sqrt{(5.6)^2 + (4.0)^2 + (5.0)^2 + (1.6)^2} = 8.7\%.$$



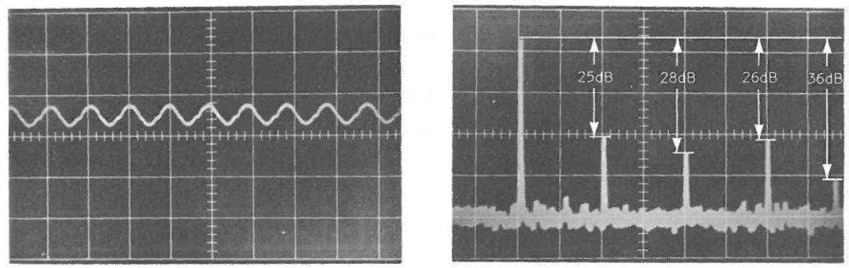
(A) TIME DOMAIN (B) FREQUENCY DOMAIN

Fig. 10-5. Zener noise source. Bias setting gives unflat frequency distribution.



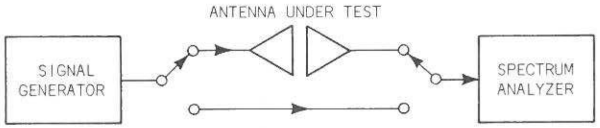
(A) TIME DOMAIN (B) FREQUENCY DOMAIN

Fig. 10-6. Zener noise source. Bias setting gives flat frequency distribution.

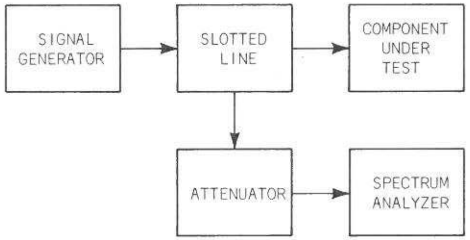


(A) TIME DOMAIN (B) FREQUENCY DOMAIN
VERTICAL - LOG, 10dB/cm

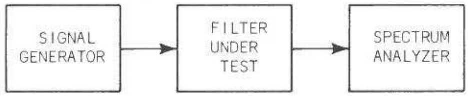
Fig. 10-7. Distortion measurement.



(A) ANTENNA TESTING

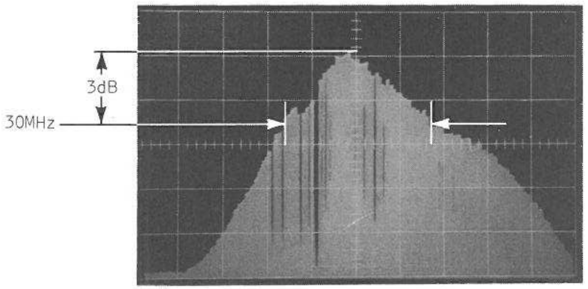


(B) SWR MEASUREMENT



(C) FILTER MEASUREMENT

Fig. 10-8. Test setups.



HORIZONTAL - 10MHz/cm
CENTER - 2GHz
VERTICAL - LIN

Fig. 10-9. Filter bandpass characteristic.

COMPONENT TRANSFER-CHARACTERISTIC MEASUREMENTS

Basically, a spectrum analyzer is a receiver. As such, it can be used in all applications where a receiver is called for. Thus, the spectrum analyzer can be used as the indicator in VSWR measurements, attenuator-insertion-loss determination, antenna pattern monitoring, filter bandpass and Q determination, etc.

The basic test setup consists of a signal generator, a spectrum analyzer and the component under test. Typical test setups are shown in Fig. 10-8. Fig. 10-9 illustrates a filter bandpass characteristic and Q measurement. The display was generated by tuning the signal generator frequency through the spectrum-analyzer dispersion, with the spectrum-analyzer plug-in in a storage oscilloscope. The filter response was thus stored on the CRT and then photographed. When storage is not available, the filter response can be obtained by controlling the shutter speed manually while the signal generator is tuned. The filter characteristic can also be obtained by connecting the spectrum-analyzer recorder output to a paper chart recorder. The filter, in this measurement, has a sufficiently high Q to permit the display of the complete response curve within one dispersion frequency width. When this is not possible, the filter characteristic is obtained by tuning the spectrum-analyzer center frequency to maintain the signal on-screen as the signal generator is tuned. The deflection amplitudes and frequencies are recorded for manual plotting of the filter response.

filter characteristic

Fig. 10-9 shows the filter characteristic taken with the spectrum-analyzer vertical control in the linear mode. The 3-dB bandwidth is measured at the 0.707 points, which at 10 MHz/cm corresponds to 30 MHz. Since $Q = f_0/\Delta f$, the loaded Q of this filter is $2000/30 \cong 67$.

SYNCHRONIZED SWEEPER TECHNIQUES

In a swept front-end spectrum analyzer, such as Tektronix Type 1L5 or Type 3L5, the sweeping-oscillator frequency setting not only controls the dispersion but also uniquely determines the center frequency. By heterodyning, filtering, and amplitude leveling the sweeping-oscillator output, it is possible to generate a signal whose frequency is equal to the instantaneous frequency to which the spectrum analyzer is tuned. Such a device is the Tektronix *Swept Frequency Converter* (SFC) for use with the Type 1L5 and Type 3L5 Spectrum Analyzer Plug-in Units.

Fig. 10-10 shows a basic block diagram of the SFC and how it relates to the block diagram of the spectrum analyzer. This system can be used to characterize the transfer function characteristics of all types of components. The test methods are the same as that used in the standard sweeper-detector-oscilloscope system. The only difference is that the spectrum analyzer acts as a synchronous detector, thereby providing several advantages over the ordinary peak-detector method. These advantages are:

1) The spectrum analyzer has very high sensitivity, thereby permitting the testing of components that cannot stand the relatively high (over 100 mV) voltages needed in the detector-oscilloscope system.

2) Synchronous detection filters out the effect of sweeper harmonics, hence, making it easy to accurately characterize components having multiple passbands.

3) The frequency converter is a relatively inexpensive component when compared to a complete sweeper.

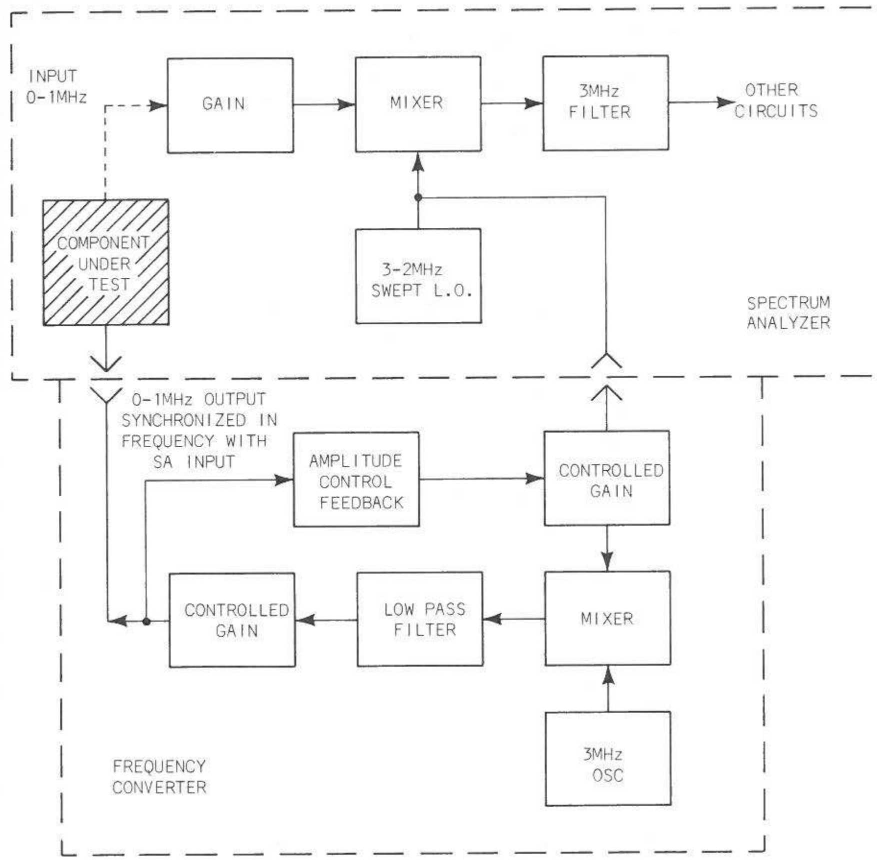


Fig. 10-10. Frequency-converter/spectrum-analyzer test system.

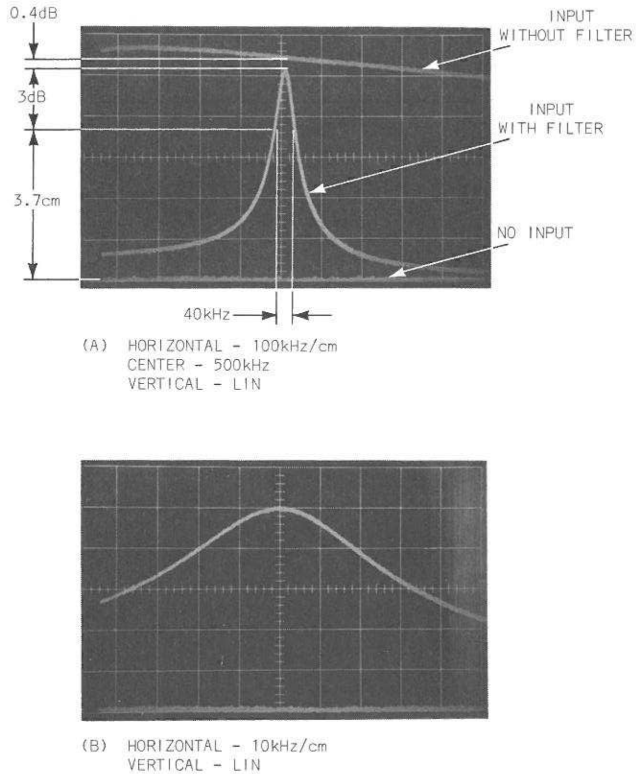


Fig. 10-11. Filter response displayed on Type 1L5 using swept frequency converter.

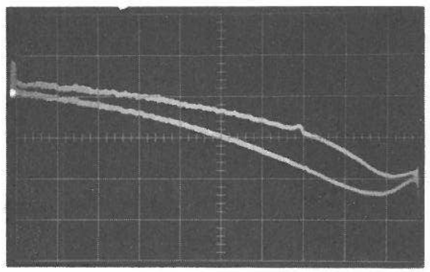


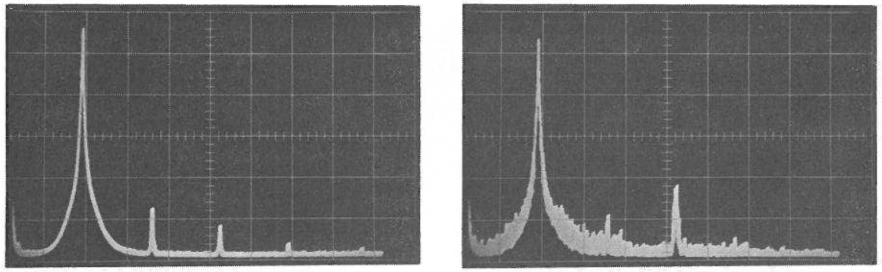
Fig. 10-12. Portable-tape-recorder frequency response.

The major application of the frequency-converter/spectrum-analyzer system is in the testing of filters. Fig. 10-11 shows the results of such a measurement. Fig. 10-11A shows the complete filter characteristic. From this display we observe that the loss in signal transfer when the filter is inserted between the SFC and the spectrum analyzer is 5.2 cm/5.4 cm, or about a 0.4-dB insertion loss.

The bandwidth at the 3-dB-down points is approximately 40 kHz. This is more accurately determined from an expanded display such as shown in Fig. 10-11B. The loaded Q of the filter can be computed as $Q = f_0/\Delta f = 500/40 = 12.5$. More detailed skirt selectivity data can be obtained by operating the spectrum analyzer in its logarithmic vertical mode. In short, as with any other sweeper-detector system, this method permits the complete characterization of a filter. One important precaution is the necessity for adequate impedance matching. Some spectrum analyzers, such as Tektronix Type 1L5 or Type 3L5, have a high impedance input. It is, therefore, necessary to use appropriate terminations or minimum loss pads to assure an impedance match between the Swept Frequency Converter, filter and spectrum analyzer.

tape-recorder performance

Besides the standard filter measurement applications, the analyzer/frequency-converter system can be used in many other areas. One illustration is in the testing of tape recorders. In this application, the output of the SFC is connected to the recorder instead of the microphone output, and the playback goes to the spectrum analyzer. This permits the measurement of such things as recorder frequency response and distortion. Fig. 10-12 shows the frequency response of a portable tape recorder at the two extremes of the tone adjustment settings. The upper trace shows a 3-dB-down point of about 30 kHz while the lower trace, at a different tone control setting, shows a 3-dB-down point of about 20 kHz. The "birdie" at about 70 kHz is caused by a bias oscillator inside the tape recorder.



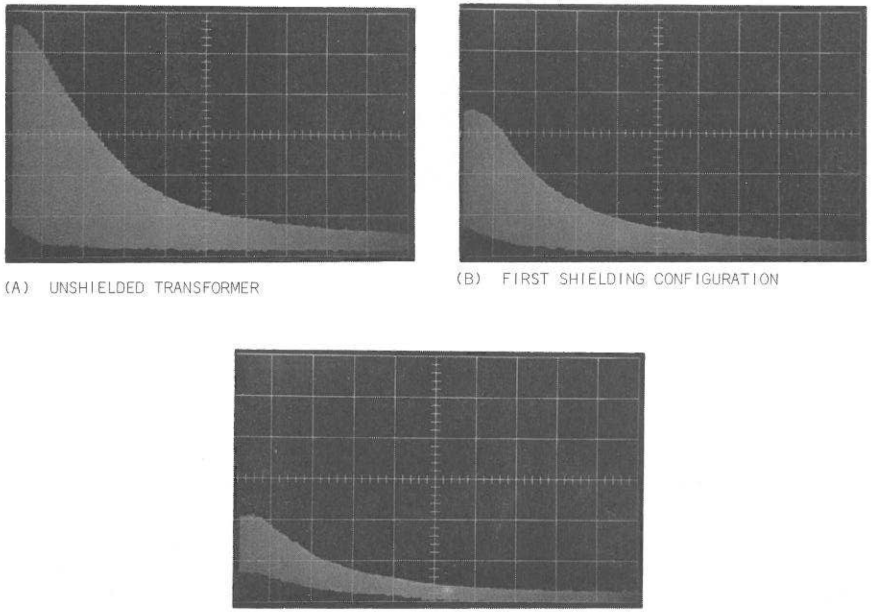
HORIZONTAL - 1kHz/cm
VERTICAL - LOG

(A) SIGNAL SOURCE

HORIZONTAL - 1kHz/cm
VERTICAL - LOG

(B) RECORDER PLAYBACK

Fig. 10-13. Portable-tape-recorder distortion measurement.



(A) UNSHIELDED TRANSFORMER

(B) FIRST SHIELDING CONFIGURATION

(C) SECOND SHIELDING CONFIGURATION

Fig. 10-14. Shielding-effectiveness test on power transformer. Horizontal – 200 Hz/cm, center – 1 kHz, and vertical – LIN.

Fig. 10-13 illustrates the measurement of the distortion characteristics of the tape recorder. Here, a 2-kHz tone was checked for harmonic content straight out of the signal generator (Fig. 10-12A) and as a playback from the tape recorder (Fig. 10-12B). It will be observed that going through the recorder has increased the third harmonic by about 10-dB.

EMI MEASUREMENTS

Spectrum analyzers make excellent electromagnetic-interference monitoring tools. Workers in the field have found good correlation in the measurements of absolute levels of interference using spectrum analyzers[3]. The determination of absolute levels of RFI and spectrum signature emanations is a complex subject requiring a discussion of the various existing specifications and standards. This is beyond the scope of this volume. There are also many applications which simply require a relative amplitude determination. This is illustrated in Fig. 10-14. Fig. 10-14A shows the line frequency interference generated by a power transformer as picked up by a single-turn magnetic loop antenna and displayed on a Tektronix Type 1L5 Spectrum Analyzer. The object is to investigate various shielding configurations. Clearly, the shielding configuration resulting in Fig. 10-14C is best. The level of interference voltage has been reduced by a factor of 5.5 cm/2 cm = 2.75 times, or 8.8 dB.

[3] Metcalf, et al., "Investigation of Spectrum Signature Instrumentation," *IEEE Trans*, EMC-7, No. 2, June, 1965.

TELEMETRY SUBCARRIER TESTS

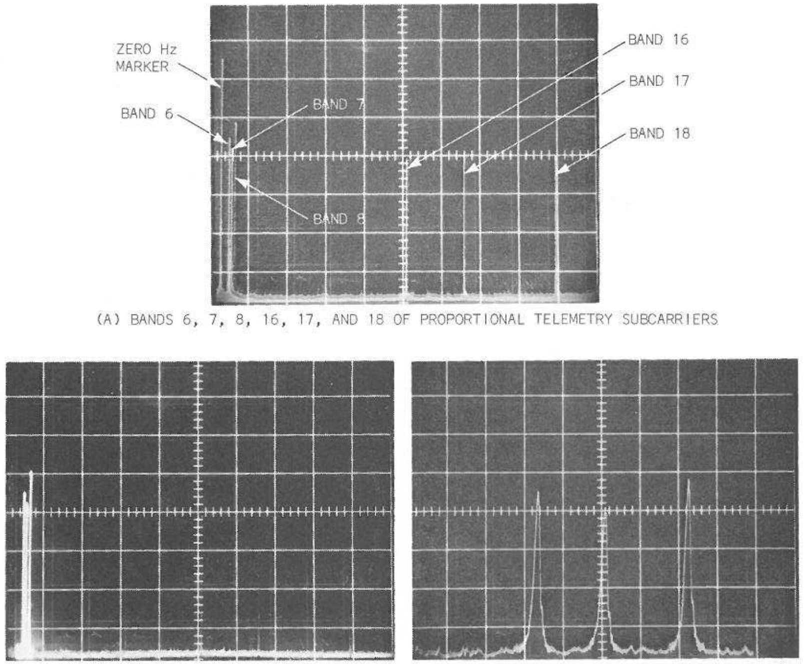
Many sets of telemetry data are usually transmitted together by modulating a set of subcarriers which are then used to modulate the main carrier by multiplexing. Subcarrier bands can have either a constant frequency spacing or a proportional frequency spacing. In the proportionally spaced system, the lower-frequency bands are spaced closer in frequency than the upper-frequency bands. The proportional spacing scheme is illustrated by the table of center frequencies, Table 10-1.

Frequently it is desired to check on the frequency spacing, amplitude, and presence or absence of these subcarriers. The spectrum analyzer does an excellent job here, except for the difficulty caused by the crowding of the lower-frequency channels. One solution is to use a specialized logarithmically sweeping spectrum analyzer to spread out the CRT spacing of the lower-frequency bands. Another solution is to observe the complete set of bands and then expand the region of interest across the full CRT. The latter technique is described in the following.

Fig. 10-15A shows bands 6, 7 & 8 and 16, 17 & 18 as displayed using a Tektronix Type 3L5 Spectrum Analyzer, Type 3B3 delayed-sweep Time Base and Type 564B Storage Oscilloscope. Note that the frequency spacing between the early bands is very close so that a detailed examination is difficult. Fig. 10-15B was taken under identical spectrum-analyzer settings but the time base was operated in the intensified mode with the intensification delay time adjusted to cover only the desired early bands. In Fig. 10-15C, the spectrum analyzer settings remain unchanged but the time base has been switched into the delayed mode of operation. This expands the intensified portion of the trace across the full width of the CRT so that a detailed examination can be easily made.

BAND NUMBER	CENTER FREQUENCY (Hz)
1	400
2	560
3	730
4	960
5	1300
6	1700
7	2300
8	3000
9	3900
10	5400
11	7300
12	10500
13	14500
14	22000
15	30000
16	40000
17	52500
18	70000

Table 10-1. IRIG proportional subcarrier bands.



(A) BANDS 6, 7, 8, 16, 17, AND 18 OF PROPORTIONAL TELEMETRY SUBCARRIERS

(B) BANDS 6, 7, AND 8 INTENSIFIED

(C) INTENSIFIED PORTION EXPANDED

Fig. 10-15. Checking proportional-spacing telemetry subcarriers by expansion method.

SIGNAL PURITY MEASUREMENTS

A multitude of sins having names such as stability, drift, incidental FM, noise sidebands, etc., can be lumped under the general heading of signal purity. Many of these names simply reflect a different way of looking at or measuring the same thing. Here, it is not the intent to give a lengthy treatise on this subject. Those interested are referred to the Bibliography. The use of spectrum analyzers to make signal-purity, or stability, measurements is illustrated by the following specific examples.

drift

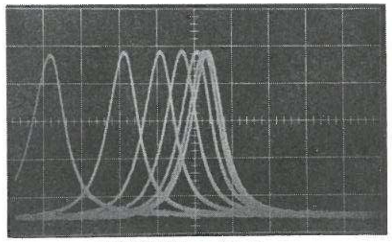
Drift is essentially a long-term phenomenon. How long a time interval is needed for "long term" is a difficult question to answer – the reader is referred to the literature for various interpretations[4].

For our purposes, long term means basically a longer interval than the measurement time of the spectrum analyzer.

Drift simply means a frequency change. The amount of change as a function of time or oscillator temperature, or some other parameter, is easily recorded by checking periodically on a spectrum analyzer. Of course, it goes without saying, that the spectrum-analyzer drift performance must be considerably better than the amount to be measured.

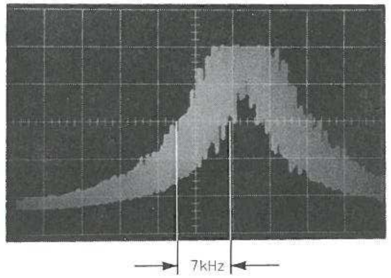
Fig. 10-16 shows a drift measurement. Oscillator basic frequency is 100 kHz, the spectrum analyzer was set at a dispersion of 1 kHz/cm. The spectrum analyzer was actuated on single sweep in one minute intervals and the resulting spectra recorded on a storage CRT. From the photograph, we observe that the total drift was about 4 kHz, initial drift was about 2 kHz/min, and the oscillator settled down after five or six minutes. Depending on the cause of this drift, we can now specify parts per million, change per degree temperature or change per volt in the power-supply voltage or whatever.

[4] Kauffman & Engelson, "Frequency Domain Stability Measurements," *The Microwave Journal*, May, 1967.



HORIZONTAL - 1kHz/cm
CENTER - 100kHz
VERTICAL - LIN

Fig. 10-16. Frequency drift measurement.



HORIZONTAL - 5kHz/cm
CENTER - 4GHz
VERTICAL - LIN

Fig. 10-17. Unresolved incidental frequency modulation.

incidental FM

Incidental frequency modulation is basically a short-term phenomenon, where the frequency change of interest occurs in less than the measurement time. Incidental FM can be at a random rate, such as caused by power-supply noise, random vibration or other random phenomena; or it can be coherent, such as caused by power-supply ripple at line frequency. Both random and coherent FM look basically alike on the spectrum analyzer when the spectrum-analyzer resolution bandwidth is wider than the FM rate. Under these conditions, the FM sidebands are not resolved and the spectrum analyzer shows a "smearing" of the response characteristic as illustrated in Fig. 10-17. Here the FM was caused by 120-Hz power-supply ripple, but this is not apparent from the display. What is apparent is that the peak-to-peak FM deviation is about 7 kHz.

From this one can compute the desired stability data in parts per million as:

$$\frac{\text{peak-to-peak deviation}}{\text{center frequency}}10^6 = \frac{7}{4000}10^6 = 1750 \text{ P/M}$$

Sometimes the incidental-FM rate is such that the individual sidebands can be resolved by the spectrum analyzer. This type of display was already discussed in Chapter 8 under Narrowband FM. Refer to that discussion for details.

Sometimes the signal impurity is due to AM rather than FM modulation. This can be described in terms of the standard AM sidebands. Describing the instability in terms of standard sidebands, whether AM or FM, becomes difficult when the modulation is random, such as due to noise. A useful technique, here, is to designate that the noise sideband power is so many dB down from the carrier, when measured with a stated-noise-bandwidth amplifier at a given frequency distance removed from the carrier. Fig. 10-18 illustrates this type of measurement.

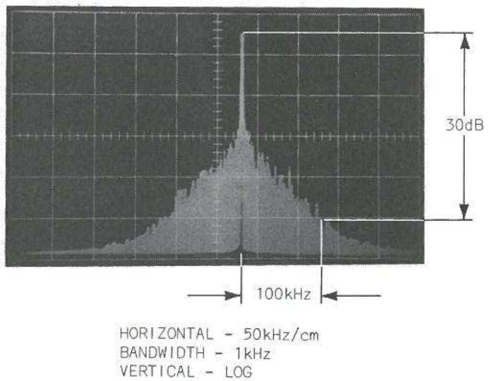


Fig. 10-18. AM noise modulation measurement.

Here one could, for example, specify that the noise sideband power, in a 1-kHz bandwidth, is 30 dB below the carrier power, 100 kHz removed from the carrier frequency. Since spectrum analyzers do not usually have specified resolution noise bandwidths, this generally has to be either calculated or measured. The determination simply consists of calculating the average width of the resolution bandwidth power (square law) curve.

DOPPLER VELOCITY MEASUREMENT

The Doppler principle permits the measurement of velocity by bouncing electromagnetic energy off a moving target and measuring the frequency difference between the transmitted and reflected radiation. The best known Doppler velocity measuring system is Doppler radar. For a CW system, the Doppler difference frequency (f_D) is related to the transmitted frequency (f_t), the target velocity (v) and the speed of electromagnetic radiation (c), by:

$$f_D \cong \frac{2v}{c} f_t.$$

For a 100-MHz radar frequency, this represents a Doppler frequency of 29.3 Hz for a 100-mile-per-hour target. These representative numbers indicate that low-frequency spectrum analyzers, such as the Tektronix Type 1L5 or Type 3L5, are well suited to this application. Depending on the basic radar frequency, the horizontal scale of the analyzer can be calibrated directly in miles per hour. The target velocity is then read from the horizontal position of the signal on the screen.

A more complex Doppler velocity measurement is illustrated by Fig. 10-18. Here, the spectrum analyzer is used to measure the Doppler frequency of laser radiation scattered by a moving fluid.

The coherent light of a laser beam is split into two beams of equal path length. The beam directions are so arranged that one beam impinges directly onto a photomultiplier tube while the other beam is pointed away from the photomultiplier, as shown in Fig. 10-19. Some of the light from the main beam, which is pointed away from the photomultiplier, is scattered by the moving fluid and enters the photomultiplier along with the light from the other beam. The photomultiplier acts as a mixer, combining the two light beams and producing an electrical signal at the difference frequency. This difference frequency is determined by the velocity of the fluid, which is, thus, indirectly measurable by the spectrum analyzer.

USING TRANSDUCERS

Spectrum analyzers can be used as frequency-selective instruments for signals involving transducers. These can involve such diverse applications as cyclical temperature effects, resonant-frequency determination for vibrating bodies, frequency and amount of displacement for various structures, pressure gradients or changes of moving fluids, etc.

Following photographs illustrate the kind of results that one can get.

Fig. 10-20A is a time-domain display of the vibration of the floor in a large industrial building under normal usage. The pickup is a velocity transducer having a sensitivity of 600 mV for every inch per second of velocity. From this display, we observe that the basic frequency is about 1/0.1 = 10 Hz and that the peak-to-peak velocity excursion is $200 \cdot 10^{-6}/600 \cdot 10^{-3} =$ 333 μin/s. Fig. 10-20B shows the frequency-domain characteristics associated with Fig. 10-20A. The frequency of greatest velocity is clearly closer to 7 Hz than the 10 Hz estimated from time-domain data. The transducer output at this frequency is 400 mV RMS. The floor velocity at the 7-Hz resonant frequency is $400/600 \cdot 10^3 = 667$ μin/s RMS, or $(667)(2.8) \cong 1870$ μin/s P-P. The spectrum-analyzer data shows that, at the resonant frequency, the floor is moving a great deal more than the average time-domain data indicates.

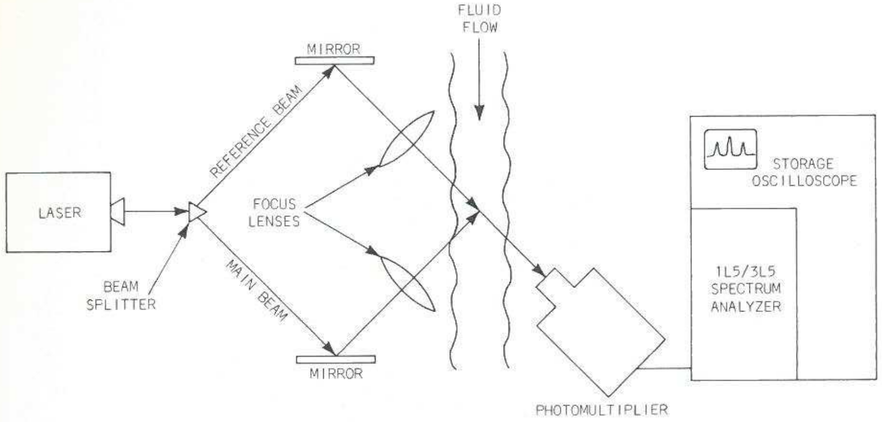


Fig. 10-19. Doppler method of fluid velocity measurement.

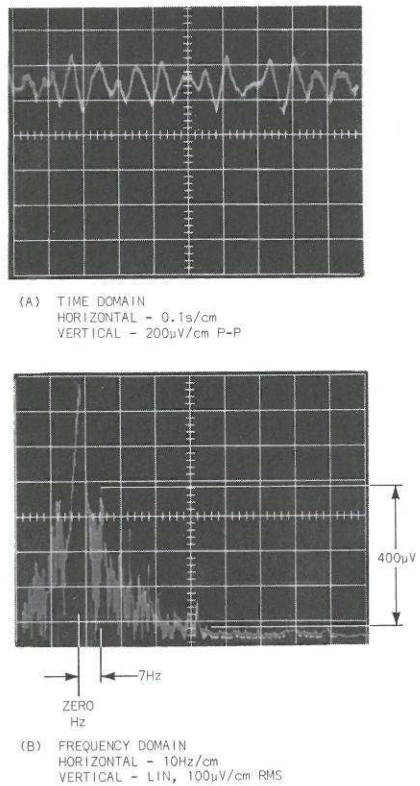
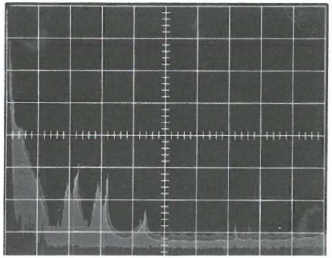


Fig. 10-20. Floor movement – velocity transducer, 600 mV for 1-in/s velocity.



HORIZONTAL - 10kHz/cm
CENTER - 50kHz
VERTICAL - LOG

Fig. 10-21. Resonance effects of small mechanical structure as picked up by vibration transducer and displayed on Tektronix Type 3L5 Spectrum Analyzer.

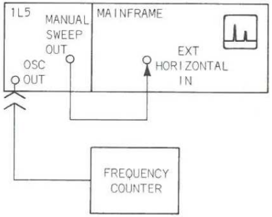
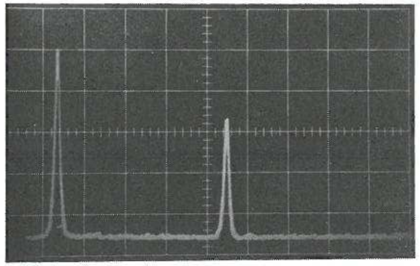


Fig. 10-22. Equipment arrangement for manual-sweep operation.

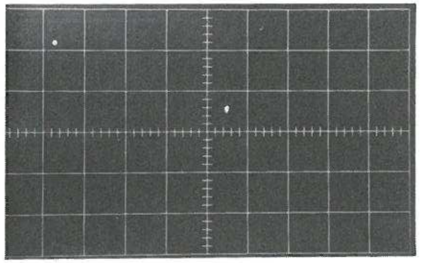
Fig. 10-21 shows another transducer application. Here, a small mechanical structure is made to vibrate by excitation from a loudspeaker type of exciter which was in turn driven by a low-frequency squarewave. The pickup is a velocity transducer. Clearly, most of the output is at low frequencies. This should be expected, since the driving source (squarewave) has most of its output at the low end of the spectrum. While the low-frequency effects are mainly due to the driving waveform, the peaks at 20 kHz, 30 kHz, and 45 kHz are resonance effects in the structure under test.

USING MANUAL SWEEP

Some swept front-end spectrum analyzers have a *manual-sweep* position, where the frequency of the swept local oscillator can be manually controlled by the operator. This, when combined with a local-oscillator front-panel output for external measurement, permits highly accurate absolute and relative frequency measurements. The basic equipment arrangement for such a measurement, using the Tektronix Type 1L5, is shown in Fig. 10-22. The local oscillator is manually tuned to the peak of the displayed signal components at which time the local-oscillator frequency is measured on the counter. The absolute frequency of each signal component can be computed by subtracting the measured local-oscillator frequency from the known IF-amplifier center frequency. Accurate frequency differences between various components are obtained by computing the difference frequencies between the various counter readings. Fig. 10-23 illustrates a typical measurement.



(A) STANDARD SPECTRUM-ANALYZER DISPLAY



(B) MANUAL SWEEP TUNED TO EACH SIGNAL

Fig. 10-23. Measuring frequency in manual sweep. Horizontal is 1 kHz/cm.

Fig. 10-23A is a standard spectrum-analyzer display of two sinusoids. Based on the spectrum-analyzer dispersion setting, the frequency difference between the two signals is 4 kHz. However, this number is only known within the accuracy of the dispersion setting, which is on the order of 5% — that is, a possible error of 200 Hz. Fig. 10-23B is a double exposure as the trace is moved to the peak of the response due to each signal. The local-oscillator output frequency was measured as 2,537,21X and 2,533,12X (the last digit is shown as X since the accuracy of the measurement makes this figure meaningless). The more accurate difference frequency is 4090 ±10 Hz rather than 4000 ±200 Hz.

The utility of the manual sweep is that the frequency of complex signals or very low-amplitude signals, which cannot be directly measured by a counter, can be accurately determined.

11

DEFINITIONS OF TERMS

In spectrum analyzers, as in other technical areas, there are a great many specialized technical terms. Many people are unfamiliar with the meaning of some of these terms, thus, creating a communications problem. The problem is compounded by the fact that different manufacturers sometimes use different terms to denote the same parameters and, conversely, the same word may have a different meaning. This problem has been recognized by the IEEE, which has established a committee whose purpose is to develop a unified set of terms and definitions. However, the task of this group will not be finished for several years.

This chapter is divided into three sections. The first section gives the terms and definitions presently used by Tektronix, Inc. The second section details the terms and definitions so far considered by the IEEE subcommittee on spectrum analyzers. The purpose of the last section is to illustrate and explain some of the more difficult terms and to indicate some of the measurement techniques used.

TEKTRONIX DEFINITIONS

Spectrum Analyzer – A device which displays a graph of relative power distribution as a function of frequency, typically on a cathode-ray tube or chart recorder.

1) *Real Time* – A spectrum analyzer that performs a continuous analysis of the incoming signal with the time sequence of events preserved between input and output.

2) *Nonreal Time* – A spectrum analyzer that performs an analysis of a repetitive event by a sampling process.
   a) *swept front-end spectrum analyzer* – A superheterodyne spectrum analyzer in which the first local oscillator is swept.

b) *swept intermediate-frequency spectrum analyzer* – A superheterodyne spectrum analyzer in which a local oscillator other than the first is swept.

Center Frequency (radio frequency or intermediate frequency) – That frequency which corresponds to the center of the reference coordinate (in units of Hz).

Center-frequency Range (radio frequency) – That range of frequency that can be displayed at the center of the reference coordinate. When referred to a control (e.g., Intermediate-frequency Center-frequency Range), the term indicates the amount of frequency change available with the control (in units of Hz).

Deflection Factor – The ratio of the input signal amplitude to the resultant displacement of the indicating spot (e.g., RMS V/div).

Dispersion (sweep width) – The frequency sweep excursion over the frequency axis of the display. Can be expressed as frequency/full frequency axis or frequency/div in a linear display.

Display Flatness – Uniformity of amplitude response over the rated maximum dispersion (usually in units of dB).

Drift (frequency drift) (stability) – Long-term frequency changes or instabilities caused by frequency changes in the spectrum-analyzer local oscillators. Drift limits the time interval that a spectrum analyzer can be used without retuning or resetting the front-panel controls (units may be Hz/s, Hz°C, etc.).

Dynamic Range (on screen) – The maximum ratio of signal amplitudes that can be simultaneously observed within the graticule (usually in units of dB).

Dynamic Range, Maximum Useful – The ratio between the maximum input power and the spectrum-analyzer sensitivity (usually in units of dB).

Frequency Band – A range of frequencies that can be covered without switching (in units of Hz).

Frequency Scale – The range of frequencies that can be read on one line of the frequency indicating dial (in units of Hz).

Incidental Frequency Modulation (residual frequency modulation) – Short-term frequency jitter or undesired frequency deviation caused by instabilities in the spectrum-analyzer local oscillators. Incidental frequency modulation limits the usable resolution and dispersion (in units of Hz).

Incremental Linearity – A term used to describe local aberrations seen as nonlinearities for narrow dispersions.

Linearity (dispersion linearity) – Measure of the comparison of frequency across the dispersion to a straight-line frequency change. Measured by displaying a quantity of equally spaced (in frequency) frequency markers across the dispersion and observing the positional deviation of the markers from an idealized sweep as measured against a linear graticule. Linearity is within $\frac{\Delta w}{w} \cdot 100\%$, where Δw is maximum positional deviation and w is the full graticule width.

Maximum Input Power – The upper level of input power that the spectrum analyzer can accommodate without degradation in performance (e.g., spurious responses and signal compression) (usually in units of dBm).

Maximum Sensitivity –

1) *Signal equals noise* – That input signal level (usually in dBm) which results in a display where the signal level above the residual noise is equal to the residual noise level above the baseline; expressed as: signal + noise = twice the noise.

2) *Minimum discernible signal* – That input signal level (usually in dBm) which results in a display where the signal is just distinguishable from the noise.

Minimum Usable Dispersion – The narrowest dispersion obtainable for meaningful analysis. Defined as ten times the incidental frequency modulation when limited by "incidental frequency modulation" (in units of Hz).

Optimum Resolution – The best resolution obtainable for a given dispersion and a given sweep time (in units of Hz). Theoretically,

$$\text{Optimum Resolution} = \sqrt{\frac{\text{dispersion (in Hz)}}{\text{sweep time (in seconds)}}}$$

Optimum Resolution (bandwidth) – The bandwidth at which best resolution is obtained for a given dispersion and a given sweep time (in units of Hz):

$$\text{Optimum Resolution (bandwidth)} = 0.66\sqrt{\frac{\text{dispersion}}{\text{sweep time}}}$$

Resolution – The ability of the spectrum analyzer to discretely display adjacent signal frequencies. The measure of resolution is the frequency separation of two equal amplitude signals, the displays of which merge at the 3-dB-down points (in units of Hz). The resolution of a given display depends on three factors: sweep time, dispersion and the bandwidth of the most selective amplifier. The 6-dB bandwidth of the most selective amplifier (when Gaussian) is called resolution bandwidth and is the narrowest bandwidth that can be displayed as dispersion and sweep time are varied. At very long sweep times, resolution and resolution bandwidth are synonymous.

Resolution (bandwidth) – Refer to resolution.

Safe Power Level – The upper level of input power that the spectrum analyzer can accommodate without physical damage (usually in units of dBm).

Scanning Velocity – Product of dispersion and sweep repetition rate (in units of Hz/unit time).

Sensitivity – Rating factor of spectrum analyzer's ability to display weak signals.

Skirt Selectivity – A measure of the resolution capability of the spectrum analyzer when displaying signals of unequal amplitude. A unit of measure would be the bandwidth at some level below the 6-dB-down points, (e.g., 10, 20, 40-dB down)(in units of Hz).

Spurious Response (spurii, spur) – A characteristic of a spectrum analyzer wherein displays appear which do not conform to the calibration of the radio frequency dial. Spurii and spur are the colloquialisms used to mean spurious responses (plural) and spurious response (singular) respectively. Spurious responses are of the following types:

1) *Intermediate-frequency feedthrough* – Wherein signals within the intermediate-frequency passband of the spectrum analyzer reach the intermediate-frequency amplifier and produce displays on the cathode-ray tube that are not tunable with the radio-frequency center-frequency controls. These signals do not enter into a conversion process in the first mixer and are not affected by the first local-oscillator frequency.

2) *Image responses* – When the input signal is above or below the local-oscillator frequency by the intermediate frequency, the superheterodyne process results in two major responses separated from each other by twice the intermediate frequency. The spectrum analyzer is usually calibrated for only one of these responses. The other is called the image.

3) *Harmonic conversion* – The spectrum analyzer will respond to signals that mix with harmonics of the local oscillator and produce the intermediate frequency. Most spectrum analyzers have dials calibrated for some of these higher-order conversions. The uncalibrated conversions are spurious responses.

4) *Intermodulation* – In the case of more than one input signal, the myriad of combinations of the sums and differences of these signals between themselves and their multiples creates extraneous responses known as intermodulation. The most harmful intermodulation is third order, caused by the second harmonic of one signal combining with the fundamental of another.

5) *Video detection* – The first mixer will act as a video detector if sufficient input signal is applied. A narrow pulse may have sufficient energy at the intermediate frequency to show up as intermediate frequency feedthrough.

6) *Internal* – A display shown on the cathode-ray tube caused by a source or sources within the spectrum analyzer itself and with no external input signal. Zero frequency feedthrough is an example of such a spurious response.

7) *Anamalous IF responses* – The filter characteristic of the resolution-determining amplifier may exhibit extraneous passbands. This results in extraneous spectrum-analyzer responses when a signal is being analyzed.

Sweep Repetition Rate – The number of sweep excursions per unit of time, sometimes approximated as the inverse of sweep time for a free-running sweep.

Sweep Time – The time required for the spot in the reference coordinate (frequency in spectrum analyzers) to move across the full graticule width (can be expressed as time/div in a linear system).

Zero-frequency Feedthrough (zero pip) – The response of a spectrum analyzer which appears when frequency of the first local oscillator is equal to the intermediate frequency. This corresponds to zero input frequency and is sometimes deliberately not suppressed so as to act as a zero-frequency marker.

IEEE DEFINITIONS

Center Frequency – That frequency which corresponds to a linear frequency span (Hz).

Deflection Factor – The ratio of the input signal amplitude to the resultant output indication. The ratio may be in terms of volts (RMS) per division, dBm per division, watts per division or any other specified factor.

Display Flatness – The peak-to-peak variation in amplitude over a specified frequency span (dB).

Display Law – The mathematical law that defines the input-output function of the instrument.

1) *Linear* – A display in which the scale divisions are a linear function of the input voltage.

2) *Square (power)* – A display in which the scale divisions are a linear function of input power.

3) *Logarithmic* – A display in which the scale divisions are a logarithmic function of the input signal.

Frequency Range – That range of frequencies over which the instrument performance is specified (Hz to Hz).

Frequency Response – The peak-to-peak variation of the displayed amplitude over a specified center-frequency range, measured at the center frequency (dB).

Frequency Span – The magnitude of the frequency segment displayed (Hz, Hz/div).

Intesifier (Baseline Clipper) – A means for changing the relative brightness between the signal and baseline portions of the display.

Resolution (R) – The ability to display adjacent responses discretely (Hz dB down). The measure of resolution is the frequency separation of two responses which merge with a 3-dB notch.

1) *Equal-amplitude signals* – As a minimum, instruments will be specified and controls labeled on the basis of two equal amplitude responses under the best operating conditions.

2) *Unequal-amplitude signals* – The frequency difference between two signals of specified unequal amplitude when the notch formed between them is 3 dB down from the smaller signal shall be termed *Skirt Resolution.*

3) *Optimum resolution* – For every combination of frequency span and sweep time there exists a minimum obtainable value of resolution (R). This is the optimum resolution (R_o), which is defined theoretically as:

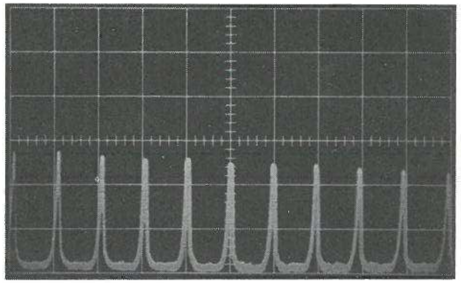
$$R_o = K\sqrt{\frac{\text{Frequency Span}}{\text{Sweep Time}}}$$

The factor "K" shall be unity unless otherwise specified.

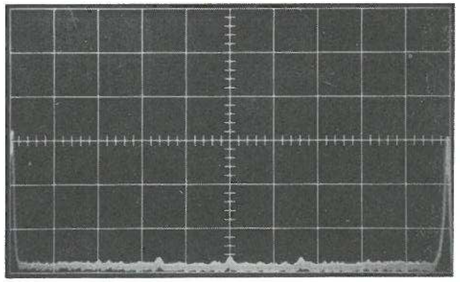
Sensitivity – Measure of a spectrum analyzer's ability to display minimum-level signals (volts, dBm). IF bandwidth, display law, and any other influencing factors must be given.

1) *Equivalent input noise* – The average level of a spectrum analyzer's internally generated noise referenced to the input.

2) *Input signal level* – The input signal level that produces an output equal to twice the value of the average noise alone. This may be a power or voltage relationship, but must be so stated.

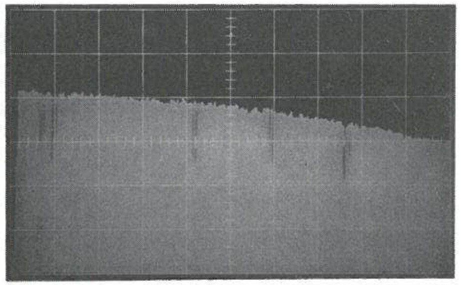


(A) INPUT MARKERS AT 10kHz FREQUENCY SPACING



(B) INPUT MARKERS AT 100kHz FREQUENCY SPACING

Fig. 11-1. Illustrating two ways of defining frequency span.



HORIZONTAL - 5MHz/cm
CENTER - 100MHz
VERTICAL - LIN

Fig. 11-2. Measuring display flatness.

EXPLANATION, MEASUREMENT METHODS

Dispersion (Frequency Span) – Fig. 11-1 illustrates the meaning and determination of the frequency span. From Fig. 11-1A we would say that the frequency span is 10 kHz/cm, while from Fig. 11-1B it follows that the frequency span is 100 kHz. With a 10-cm graticule width, the two statements are of course equivalent.

Display Flatness – Fig. 11-2 illustrates a display-flatness measurement. A constant-amplitude input signal is tuned across the dispersion with the spectrum analyzer set to the center frequency of interest. The measurement consists of noting the maximum and minimum display heights and computing the ratio in dB. For the case of Fig. 11-2, the display flatness exhibits a variation of $20 \log \frac{4.2 \text{ cm}}{3 \text{ cm}} = 2.9$ dB.

Besides specifying the dispersion or center frequency, it is sometimes important to specify the output impedance or VSWR of the generator used. Most spectrum analyzers will exhibit a degradation in display flatness when checked with a signal source of poor VSWR. When in doubt, it is best to insert a pad (about 10 dB will do) between spectrum analyzer and signal generator.

A complete statement of the results of a display-flatness measurement might be something like this: Display flatness – 2.9 dB peak-to-peak over 50-MHz dispersion at 100-MHz center frequency with a 10-dB 50-Ω pad between spectrum analyzer and signal generator.

Dynamic Range (on screen) – Among the several dynamic ranges, the on-screen dynamic range is the most frequently referred to by the user. As the definition stated, this is a number in dB indicating the ratio of the largest to the smallest input signal levels that can be observed on the CRT screen.

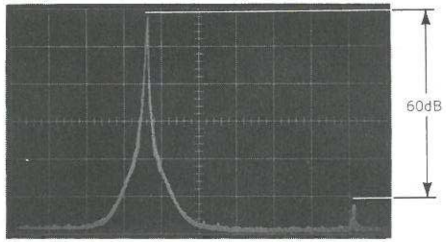
Fig. 11-3 illustrates this measurement. The basic technique calls for the simultaneous introduction of two CW signals into the spectrum analyzer. One signal amplitude is adjusted to give a full-screen deflection while the other signal amplitude is adjusted for the minimum signal level of interest. The minimum signal level may be a given number of centimeters deflection, or the sensitivity level of the instrument or some other specified level. The ratio between these two signal levels is the dynamic range on screen. This technique is illustrated by Fig. 11-3A.

When two independent signal sources are not available, the measurement is made by reducing the level of a single signal source from that which yields a full-screen deflection to the minimum desired level. The ratio of the two levels is the dynamic range on screen. This method is illustrated by the difference in the deflection levels of the large signal in Fig. 11-3A and the reduced level of the same signal shown in Fig. 11-3B.

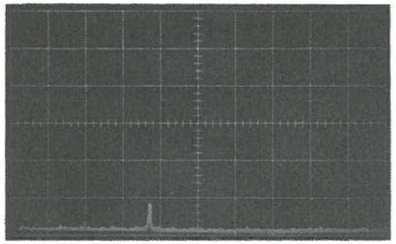
The two-signal method is preferred since it takes into account the effect of the presence of the large signal on the deflection of the small signal. Effects such as desensitization or gain compression are accounted for in the two-signal method but not in the one-signal method.

Since different display laws give a different dynamic range on screen, it is important that the display law be clearly indicated. In Fig. 11-3, the display law is logarithmic. This is the setting that corresponds to the largest dynamic range on screen. The intent of the measurements shown in Fig. 11-3 was to verify the specified 60-dB dynamic range on screen in the LOG display setting. Fig. 11-3 shows that, for the spectrum analyzer in question, a signal 60 dB below that yielding a full-screen deflection is clearly observable. This means that the instrument meets the specified performance.

Incidental FM and Drift – It is sometimes difficult to distinguish between incidental FM and drift. The former is basically a short-term phenomenon, while the latter takes considerably more time. The basic difference between drift and incidental FM is illustrated in the time-frequency diagram of Fig. 11-4. For an accurate measurement of incidental FM, it is necessary that the measurement time be greater than the period of the incidental FM excursion, but not so great that the measurement accuracy is affected by the drift.



(A) VERTICAL - LOG



(B) VERTICAL - LOG

Fig. 11-3. Dynamic-range on-screen measurement illustrating a 60-dB signal ratio.

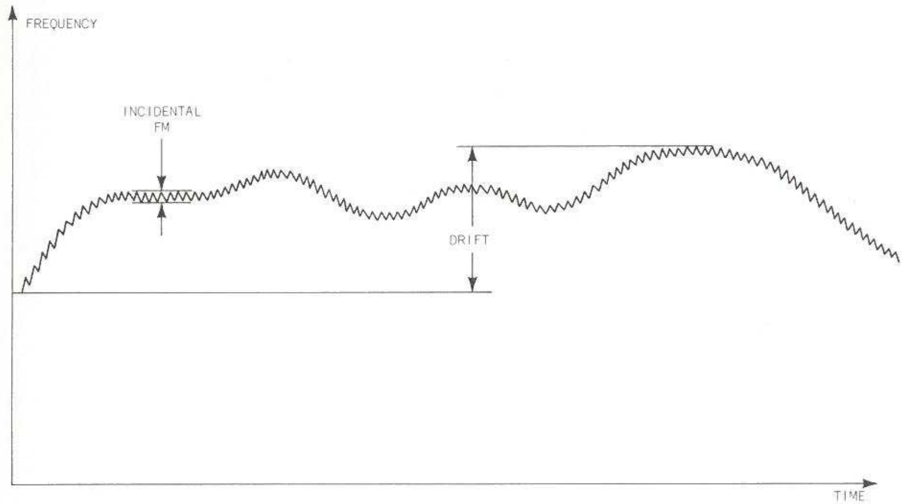


Fig. 11-4. Difference in excursion period between drift and incidental FM.

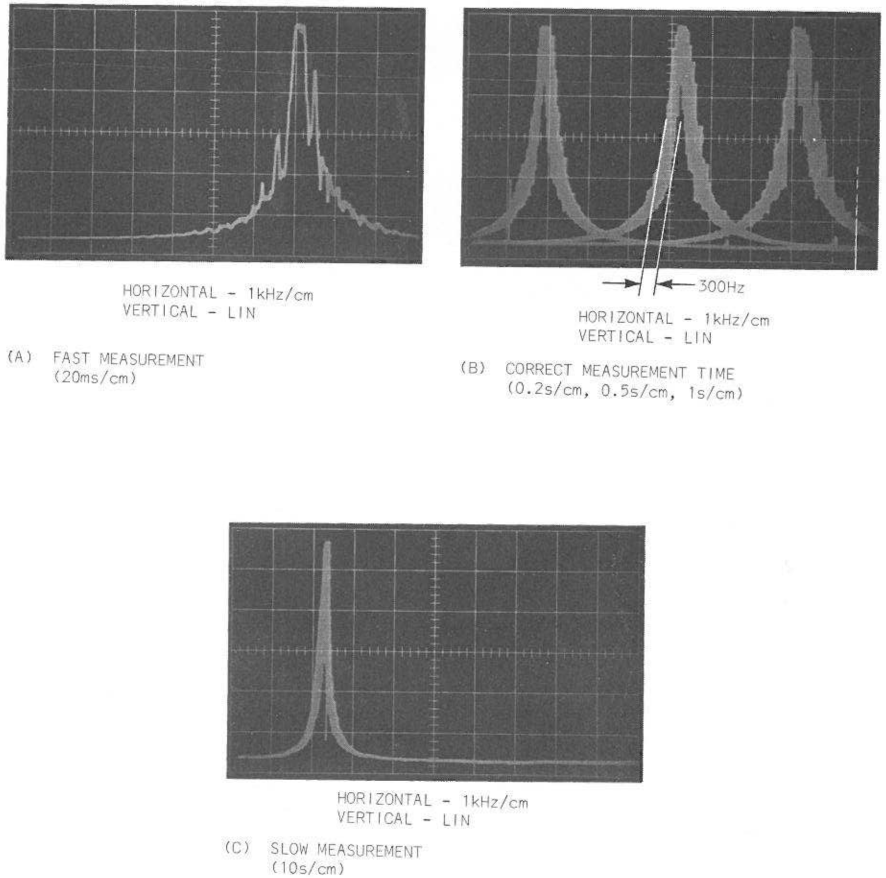


Fig. 11-5. Effect of measurement time on incidental-FM determination.

The effect of measurement time is shown in Fig. 11-5. The incidental FM in question is caused by power-supply ripple at 60 and 120 Hz. The period of the incidental FM is, therefore, on the order of 10 ms. A sweep of 20 ms/cm, as shown in Fig. 11-5A, is clearly insufficient to capture the full frequency excursion of the incidental FM. At sweeps of 200 ms/cm to 1 s/cm, as shown in Fig. 11-5B, the full frequency excursion of the incidental FM is observed. The average broadening of the trace is about 0.3 cm. This corresponds to 300 Hz of incidental FM, since the dispersion is 1 kHz/cm. Fig. 11-5C shows an incidental FM of only about 200 Hz. The apparent reduction of 100 Hz is due to the cancelling effect of the drift during the long measurement time of 10 s/cm. If the drift were in the opposite direction, the incidental FM would appear to have increased.

The drift measurement is illustrated in Fig. 11-6. Here, the signal frequency was checked at two-minute intervals. It took six such intervals for the signal to drift across the 10-kHz screen width. The average drift rate is, therefore, about 10 kHz/12 min = 834 Hz/min.

There is one important precaution when making these measurements. The stability of the signal must be considerably better than that of the spectrum analyzer. If this condition is not met, the results will show the combined effects of signal and spectrum analyzer rather than the spectrum analyzer alone.

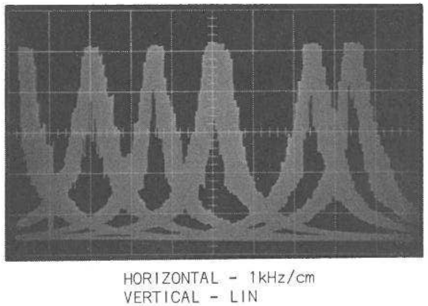


Fig. 11-6. Measuring frequency drift (single sweep every 2 minutes).

Linearity – The measurement technique is of considerable importance when specifying frequency linearity; this is because different assumptions of where the ideal straight line should be lead to results that may differ by several orders of magnitude. Fig. 11-7 illustrates several ways of drawing the hypothetical ideal straight line. As will be observed, the largest error is about twice the magnitude of the smallest.

The most popular measurement technique involves the use of ten equally spaced frequency markers. The marker spacing and center frequency are so adjusted that the second and tenth markers fall on top of a graticule line for a linear ten-division graticule. The frequency linearity error is then determined by observing the maximum positional deviation (Δw) of the other markers and comparing this to the full graticule width. This measurement technique is illustrated in Fig. 11-8. Fig. 11-8A shows a display of eleven markers covering ten equal frequency intervals. The second marker is on the second graticule line and the tenth marker is on the tenth graticule line which is nine intervals from the graticule beginning. It will be observed that the middle marker has the maximum positional deviation. To measure this deviation more accurately, the horizontal trace may be expanded, as shown in Fig. 11-8B. Here, the central marker is off by 0.5 cm at a 5 times expansion. The maximum positional deviation on the unexpanded sweep is, therefore, 0.1 cm. The linearity error can now be computed as

$$\frac{\Delta w}{w} \cdot 100 = \frac{0.1}{10} \cdot 100 = 1\%.$$

Sensitivity – When measuring sensitivity, it is important that the spectrum analyzer be properly optimized. This includes: optimization of mixer peaking, not sweeping too fast, setting the analyzer for the appropriate display mode, having the specified resolution bandwidth and setting the gain for a reasonable amount of vertical noise.

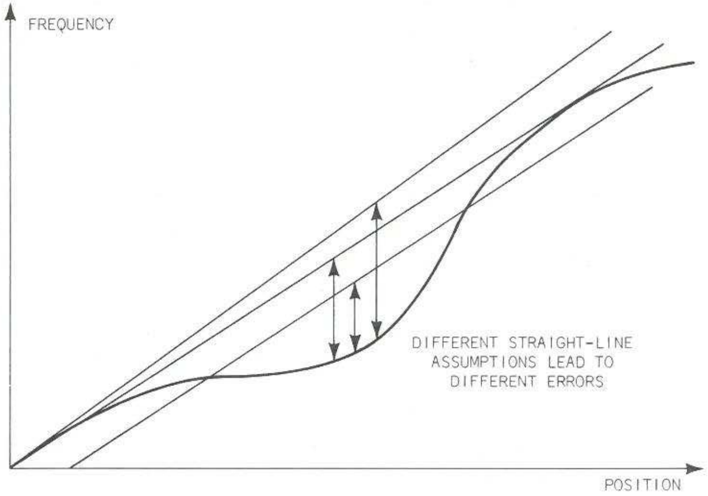
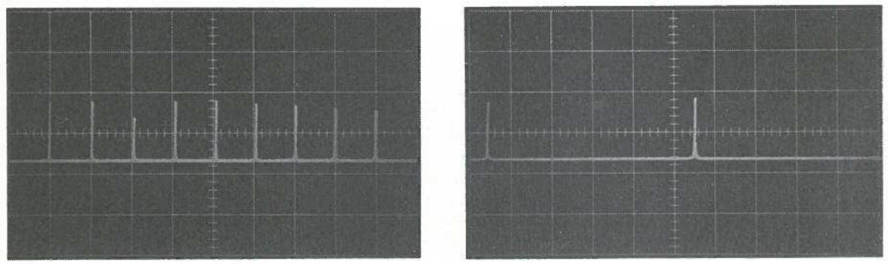


Fig. 11-7. Different ways of specifying frequency-linearity error.



(A) ELEVEN MARKERS, TEN FREQUENCY INTERVALS, MARKERS ON SECOND AND TENTH GRATICULE LINES

(B) HORIZONTAL EXPANSION OF (A) BY A FACTOR OF FIVE

Fig. 11-8. Measuring frequency linearity.

At Tektronix, sensitivity is specified as the signal input level that results in a signal-plus-noise deflection which is equal to twice the deflection due to noise alone, with the vertical mode in LIN, at a stated resolution bandwidth and all other factors optimized. Fig. 11-9 shows an $S + N = 2N$ deflection setting.

Resolution – The basic definition of resolution involves two equal-amplitude signals whose responses merge with a 3-dB notch. This definition is illustrated in Fig. 11-10A. Two equal-amplitude signals can be obtained in two ways. One is to use two separate signal generators with their outputs added in a resistive network. Another method is to use a balanced modulator to produce suppressed carrier AM where the two sidebands form the two equal-amplitude signals. Fig. 11-10A was obtained by use of a balanced modulator. Here, the two responses are 0.75 cm apart. At a dispersion of 20 kHz/cm, we have a resolution of 15 kHz.

Sometimes it is inconvenient to produce two equal-amplitude signals. One can then use the derived relationship that, for a Gaussian response shape, the 6-dB-down bandwidth gives a close approximation to the previously defined resolution. This is illustrated in Fig. 11-10B. Here, the 6-dB-down response width is about 0.65 cm, corresponding to a resolution of 13 kHz.

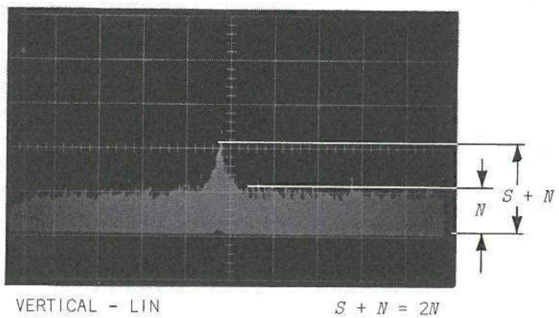
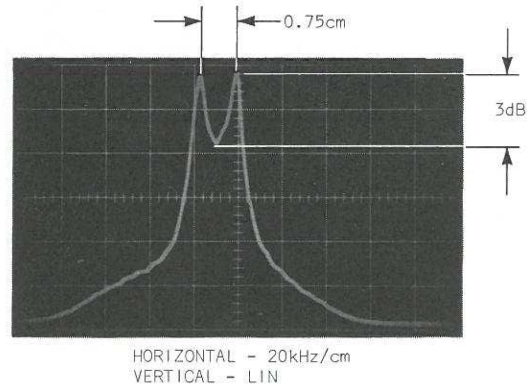
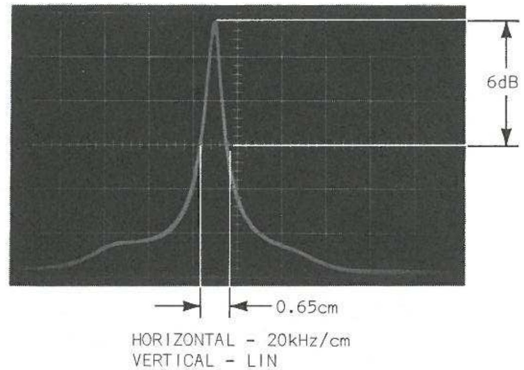


Fig. 11-9. Signal-plus-noise-equals-twice-noise type of sensitivity measurement.



(A) TWO-EQUAL-AMPLITUDE-SIGNALS METHOD



(B) SINGLE-SIGNAL 6dB-BANDWIDTH METHOD

Fig. 11-10. Measuring resolution.

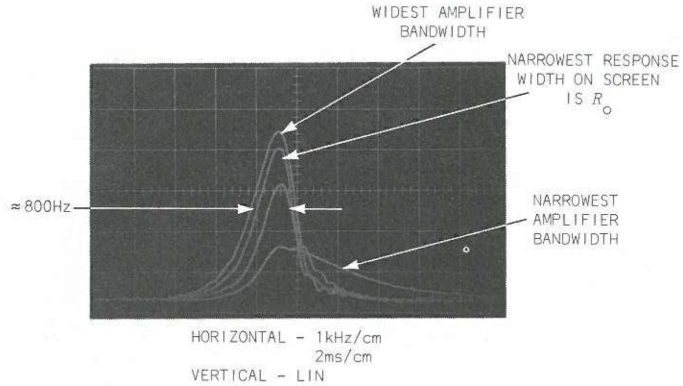


Fig. 11-11. Illustrating optimum resolution.

For every setting of sweep time and dispersion, there is some particular value of minimum obtainable resolution. Making the resolution amplifier bandwidth narrower only serves to increase the display width (resolution) observed on the screen. This minimum obtainable resolution is called the optimum resolution (R_o). The fact that an R_o exists is illustrated in Fig. 11-11. This is a multiexposure showing the changes in the display as the amplifier resolution bandwidth is changed at constant dispersion and sweep time. As the resolution bandwidth is decreased, the display amplitude decreases and the trace broadens out somewhat. Note that the narrowest and widest bandwidth settings give a wider response width than one of the intermediate settings.

The theoretical optimum resolution can be computed from

$$R_o = \sqrt{\frac{\text{dispersion}}{\text{sweep time}}}.$$

For Fig. 11-11,

$$R_o = \sqrt{\frac{1000}{2 \cdot 10^{-3}}} = \sqrt{50 \cdot 10^4} \cong 700 \text{ Hz}.$$

This is in fairly good agreement with the experimentally determined 6-dB bandwidth for the narrowest response, which is about 800 Hz.

APPENDIX

dB, dBm

The logarithmic expression of ratios is quite common in spectrum-analyzer usage. The following is a brief review of the exponential function and, its inverse, the logarithmic function, notation in *bels* and *decibels* and finally a set of tables of decibel relationships.

An exponential is a relationship of the form $a^m = P$. Some of the rules for the manipulation of exponentials are:

$$(a^m)(a^n) = a^{(m+n)} \qquad \text{(A-1)}$$

$$\frac{a^m}{a^n} = a^{(m-n)} \qquad \text{(A-2)}$$

$$a^{-m} = \frac{1}{a^m} \qquad \text{(A-3)}$$

$$a^{\frac{m}{n}} = \sqrt[n]{a^m} \qquad \text{(A-4)}$$

$$a^0 = 1 \qquad \text{(A-5)}$$

The inverse of an exponential function is a logarithmic function. Thus, the relationship $a^m = P$ can also be written as

$$\log_a P = m, \qquad \text{(A-6)}$$

where a is the base and m is the logarithm of P to the base a. Our interest is in a base of $a = 10$. This is called the System of Common Logarithms. Common logarithms are usually abbreviated as "log" – this abbreviation is supposed to indicate that the base is 10. Thus, log 100 = 2, since $10^2 = 100$.

The advantage of handling ratios in terms of logarithms is that multiplication and division are eliminated and are replaced by addition and subtraction. This follows from the relationship between exponents of exponential functions as given in equations (A-1) through (A-5). Our interest is in ratios of power. In logarithmic notation, a power ratio is given as $\log \frac{P_2}{P_1} = N$ bels. The *bel* is a dimensionless unit, simply indicating a power ratio expressed in logarithms to the base 10. The bel is a fairly large ratio, a more convenient unit is one tenth of a bel or a *decibel* (dB). There are 10 dB in every bel. A power ratio expressed in dB's is

$$10 \log \frac{P_2}{P_1} = N \text{ dB}. \tag{A-7}$$

That is – the basic power ratio in bels is multiplied by ten to get the number of decibels. In equation (A-7), P_1 is the reference power to which P_2 is compared. The result is in the form of a ratio. Thus, for $\frac{P_2}{P_1} = 100$, $N_{\text{bels}} = 2$, and $N_{\text{dB}} = 20$. Whether $P_1 = 1$ W with $P_2 = 100$ W, or $P_1 = 10$ W with $P_2 = 1$ kW makes no difference, N is 20 dB if $\frac{P_2}{P_1} = 100$.

The relationship in equation (A-7) is the basic definition for decibels, all other relationships are derived from this. Sometimes a derivation is based on assumptions or approximations that will not always hold. When in doubt, the user should go back to equation (A-7). An example is when dealing with voltage ratios across equal resistances. Thus,

$$P_1 = \frac{V_1^2}{R},$$

$$P_2 = \frac{V_2^2}{R},$$

and

$$10 \log \frac{P_2}{P_1} = 10 \log \left(\frac{V_2}{V_1}\right)^2 = N \text{ dB}$$

since the resistances cancel. Following the rules for manipulating exponents and logarithms, we get

$$10 \log \left(\frac{V_2}{V_1}\right)^2 = 20 \log \left(\frac{V_2}{V_1}\right) = N \text{ dB} \tag{A-8}$$

Equation (A-8) is only valid when dealing with a constant impedance, otherwise it is incorrect.

Sometimes expressing a specific amount of power using the dB notation is desired. This requires that the reference level P_1 be fixed. When the units are given as dBm, it means that $P_1 = 1$ milliwatt. Thus, 20 dBm means 100 mW. Frequently, a power level less than 1 mW needs to be expressed logarithmically. Here, the ratios are inverted and a minus sign added before the dBm, in accordance with the rules for exponentials, equation (A-3). Thus,

$$10 \log \frac{.001}{1 \text{ mW}} = -10 \log \frac{1000}{1 \text{ mW}} = -30 \text{ dBm}.$$

The following is a table of voltage and power ratios versus the equivalent number of dB's. The user should keep in mind that that voltage- or current-ratio method only holds true across equal impedances.

dB	VOLTAGE OR CURRENT RATIO	POWER RATIO	dB	VOLTAGE OR CURRENT RATIO	POWER RATIO
0.0	1.000	1.000	26.	19.95	398.1
0.1	1.012	1.023	27.	22.39	501.2
0.2	1.023	1.047	28.	25.12	631.0
0.3	1.035	1.072	29.	28.18	794.3
0.4	1.047	1.096	30.	31.62	1000.
0.5	1.059	1.122	31.	35.48	1259.
0.6	1.072	1.148	32.	39.81	1585.
0.8	1.096	1.202	33.	44.67	1995.
1.0	1.122	1.259	34.	50.12	2512.
1.5	1.189	1.413	35.	56.23	3162.
2.0	1.259	1.585	36.	63.10	3981.
2.5	1.334	1.778	37.	70.79	5012.
3.0	1.413	1.995	38.	79.43	6310.
4.	1.585	2.512	39.	89.13	7943.
5.	1.778	3.162	40.	100.0	10000.
6.	1.995	3.981	41.	112.2	12590.
7.	2.239	5.012	42.	125.9	15850.
8.	2.512	6.310	43.	141.3	19950.
9.	2.818	7.943	44.	158.5	25120.
10.	3.162	10.000	45.	177.8	31620.
11.	3.548	12.59	46.	199.5	39810.
12.	3.981	15.85	47.	223.9	50120.
13.	4.467	19.95	48.	251.2	63100.
14.	5.012	25.12	49.	281.8	79430.
15.	5.623	31.62	50.	316.2	100000.
16.	6.310	39.81	51.	354.8	125900.
17.	7.079	50.12	52.	398.1	158500.
18.	7.943	63.10	53.	446.7	199500.
19.	8.913	79.43	54.	501.2	251200.
20.	10.000	100.00	55.	562.3	316200.
21.	11.22	125.9	56.	631.0	398100.
22.	12.59	158.5	57.	707.9	501200.
23.	14.13	199.5	58.	794.3	631000.
24.	15.85	251.2	59.	891.3	794300.
25.	17.78	316.2	60.	1000.0	1000000.

Table A-1. Decibels.

NULL NUMBER	$t = \frac{\Delta F}{f}$
1st	2.4048
2nd	5.5201
3rd	8.6531
4th	11.7915
5th	14.9309
6th	18.0711
7th	21.2116
8th	24.3525
9th	27.4935
10th	30.6346

Table A-2. Carrier nulls, $J_0(t) = 0$.

NULL NUMBER	$t = \frac{\Delta F}{f}$
1st	3.83
2nd	7.02
3rd	10.17
4th	13.32
5th	16.47
6th	19.62
7th	22.76
8th	25.90
9th	29.05

Table A-3. First-sideband nulls, $J_1(t) = 0$.

BESSEL FUNCTIONS

Bessel functions are used extensively in frequency modulation, as discussed in Chapter 4. Our interest is restricted to Bessel functions of the first kind, integer order, and positive argument. The notation is $J_p(t)$, where: J means Bessel function of the first kind, p is the order, and t is the argument. In frequency modulation theory, the order indicates the sideband number and the argument is the modulation index ($\Delta F/f$). For more detailed tables and graphs of Bessel functions, see the references. For example, *British Association for the Advancement of Science Mathematical Tables*, University Press, Cambridge, vols. VI & X.

ARGUMENT $t = \frac{\Delta F}{f}$	CARRIER $J_0(t)$	FIRST SIDEBAND $J_1(t)$	SECOND SIDEBAND $J_2(t)$	THIRD SIDEBAND $J_3(t)$
0.0	1.000	0.000	0.000	0.000
0.2	0.990	0.100	0.005	0.0002
0.4	0.960	0.196	0.020	0.001
0.6	0.912	0.287	0.044	0.004
0.8	0.846	0.369	0.076	0.010
1.0	0.765	0.440	0.115	0.020
1.2	0.671	0.498	0.159	0.033
1.4	0.567	0.542	0.207	0.051
1.6	0.455	0.570	0.257	0.073
1.8	0.340	0.582	0.306	0.099
2.0	0.224	0.577	0.353	0.129
2.2	0.110	0.556	0.395	0.162
2.4	0.003	0.520	0.431	0.198
2.6	-0.097	0.471	0.459	0.235
2.8	-0.185	0.410	0.478	0.273
3.0	-0.260	0.339	0.486	0.309
3.2	-0.320	0.261	0.484	0.343
3.4	-0.364	0.179	0.470	0.373
3.6	-0.392	0.096	0.445	0.399
3.8	-0.403	0.013	0.409	0.418
4.0	-0.397	-0.066	0.364	0.430

Table A-4. Bessel function values.

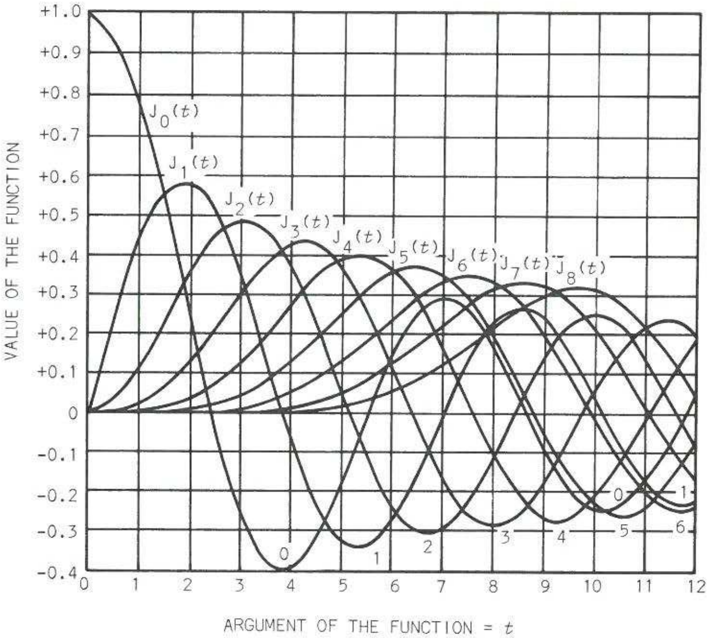
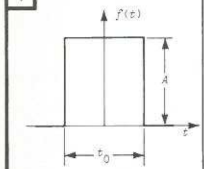
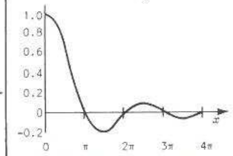
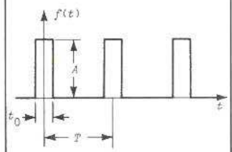
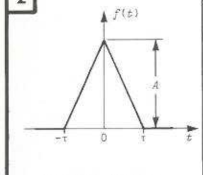
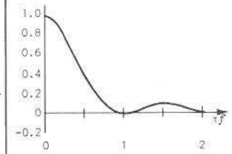
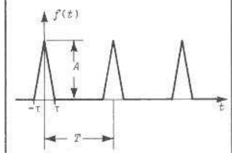
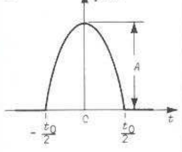
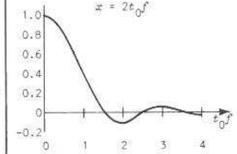
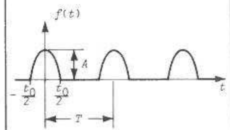
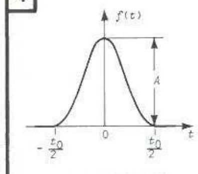
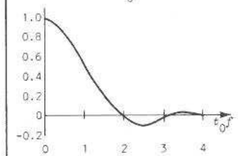
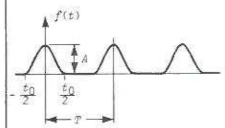
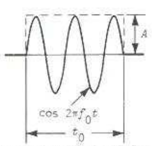
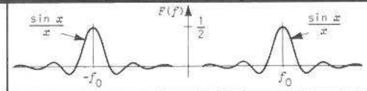


Fig. A-1. Bessel functions for the first 8 orders.

	PULSE	SPECTRUM	PULSE TRAIN	FOURIER COEFFICIENTS
1	 RECTANGLE	$F(f) = At_0 \frac{\sin x}{x}$ $x = \pi f t_0$ 		$C_n = \frac{2At_0}{T} \left\| \frac{\sin \frac{n\pi t_0}{T}}{\frac{n\pi t_0}{T}} \right\|$
2	 ISOSCELES TRIANGLE	$F(f) = A\tau \left(\frac{\sin x}{x}\right)^2$ $x = \pi f \tau$ 		$C_n = \frac{2A\tau}{T} \left(\frac{\sin \frac{n\pi\tau}{T}}{\frac{n\pi\tau}{T}}\right)^2$
3	 COSINE PULSE	$F(f) = \frac{2}{\pi} At_0 \frac{\cos \frac{\pi}{2}x}{1 - x^2}$ $x = 2t_0 f$ 		$C_n = \frac{4}{\pi} \frac{At_0}{T} \left\| \frac{\cos \frac{n\pi t_0}{T}}{1 - \left(\frac{2nt_0}{T}\right)^2} \right\|$
4	 COSINE-SQUARED PULSE	$F(f) = \frac{At_0}{2} \frac{\sin \pi x}{\pi x(1-x^2)}$ $x = t_0 f$ 		$C_n = \frac{At_0}{2T} \frac{\sin \frac{n\pi t_0}{2T}}{\frac{n\pi t_0}{2T} \left\| 1 - \left(\frac{nt_0}{2T}\right)^2 \right\|}$
5	 PULSED-RF RECTANGULAR PULSE	 $F(f) = \frac{At_0}{2}\left(\frac{\sin \pi t_0(f-f_0)}{\pi t_0(f-f_0)} + \frac{\sin \pi t_0(f+f_0)}{\pi t_0(f+f_0)}\right)$		

EXPLANATION: The fact that an RF pulse has the same shape as the pulse without the carrier, shifted by the RF frequency f_0, is a direct consequence of the frequency shift theorem, example 8 in Table 3-1. Usually, in practical spectrum analysis, the portion of the spectrum theoretically at $-f_0$ is accounted for by doubling the amplitude of its mirror image at f_0. The reasoning is the same as that when combining the two impulse functions of the infinite sinewave into one, as shown in Fig. 3-1. In theoretical spectrum analysis, however, it is helpful not to combine the two parts of the spectrum, as this permits the easy determination of the spectra of fractional-cycle sinusoidal pulses, such as that shown in number 3 of this table and discussed in the first example in Section 5, Chapter 3.

Table A-5. Fourier transforms.

FOURIER ANALYSIS

Table A-5 gives both graphical and mathematical relationships for time-domain to frequency-domain conversion.

CW SENSITIVITY

The noise power generated by a resistance is $N = kTB$, where N is noise power, k is Boltzman's constant, T is absolute temperature in degrees Kelvin, and B is the noise bandwidth. At an absolute temperature of 290°K, this is equivalent to a noise power of –114 dBm for a 1-MHz bandwidth, as discussed in Chapter 5. The actual sensitivity of an amplifier is always worse than this because of higher noise due to the active elements and because of signal losses in matching attenuators, filters or other front-end devices. The amount by which the actual sensitivity is degraded compared to the ideal sensitivity is called the noise figure. Thus, the CW sensitivity, as measured by the signal-equals-noise method, is a function of both noise figure and noise bandwidth – the temperature is usually assumed to be fixed at 290°K = 17°C.

Fig. A-2 is a graphical representation of the sensitivity, bandwidth, noise figure relationship.

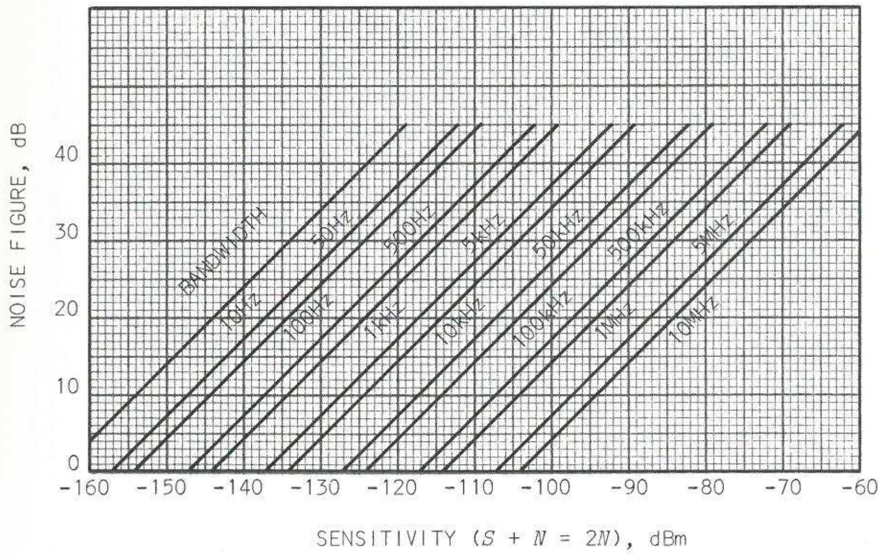


Fig. A-2. Sensitivity as a function of bandwidth and noise figure.

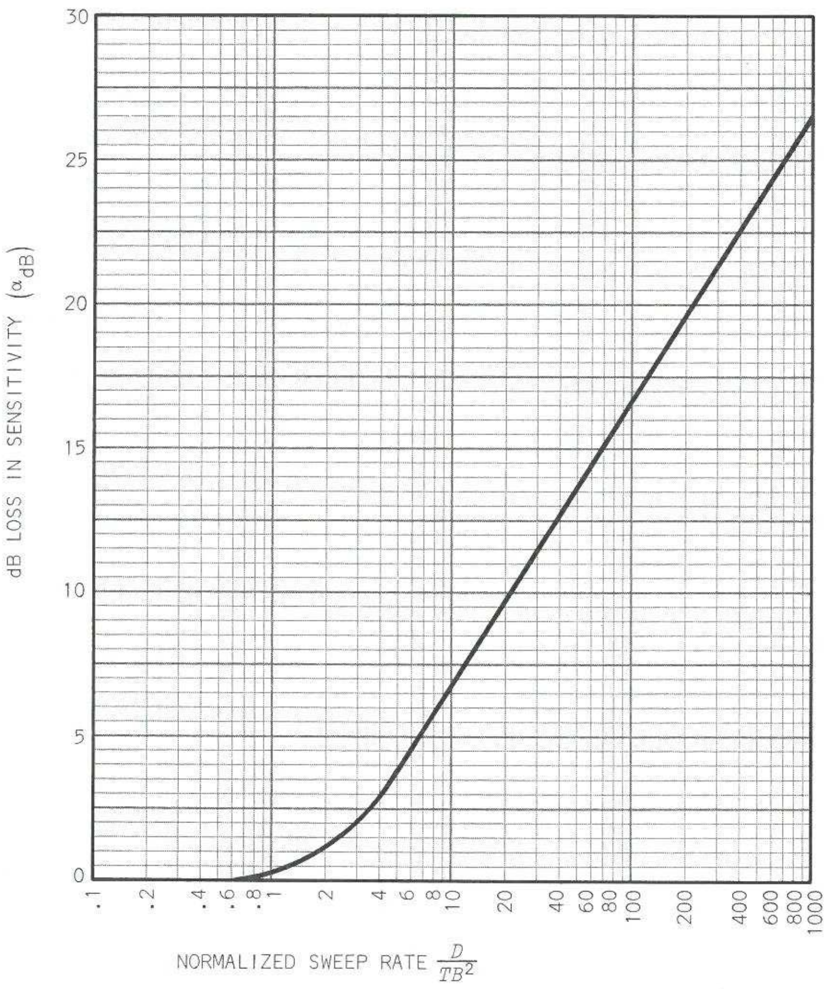


Fig. A-3. Loss in CW sensitivity as a function of normalized sweep rate. D = dispersion (Hz), B = 3 dB bandwidth (Hz), T = time (s). Based on assumption of Gaussian amplitude response.

As was indicated in Chapter 5, the CW sensitivity degrades as sweep time and/or resolution bandwidth is decreased. The equation connecting these parameters for a Gaussian resolution shape is:

$$\frac{S_o}{S} = \alpha = \left[1 + 0.195\left(\frac{D}{TB^2}\right)^2\right]^{-\frac{1}{4}}, \tag{A-9}$$

where S_o and S are the sensitivity in volts, corresponding to no degradation and degraded performance respectively. Alpha is the sensitivity ratio, a number less than unity. Thus, if $\frac{S_o}{S} = 0.1$ and $S_o = 0.1S$, we mean that it takes ten times the nondegraded (S_o) input voltage to overcome the spectrum-analyzer noise under the conditions in question. A convenient way of indicating this loss in sensitivity is in dB, which is computed:

$$\alpha_{dB} = 20 \log \alpha. \tag{A-10}$$

Since α is less than unity, α_{dB} is a negative number. However, when plotting on a graph, the negative sign is usually replaced by the word "loss," which is what the negative sign meant in the first place. Fig. A-3 is a graph of sensitivity loss as a function of full-screen dispersion D, resolution bandwidth B and full-screen sweep time T.

RESOLUTION BANDWIDTH

In Chapter 5, it was indicated that there are two terms associated with the resolution capability of the spectrum analyzer. One term is *resolution bandwidth* (B); this is the actual bandwidth of the narrowest bandwidth amplifier. A second term is *resolution* (R), which refers to the display on the CRT screen. At long sweep times, the display on the CRT is a tracing of the response characteristic of the spectrum-analyzer passband and the two terms become synonymous. At short sweep times, the display on the screen indicates a wider bandwidth than the amplifier actually has. The ratio of the screen display (R) and the true bandwidth can be computed from the relationship:

$$\frac{R}{B} = \left[1 + 0.195\left(\frac{D}{TB^2}\right)^2\right]^{\frac{1}{2}} \qquad \text{(A-11)}$$

This equation is plotted in Fig. A-4.

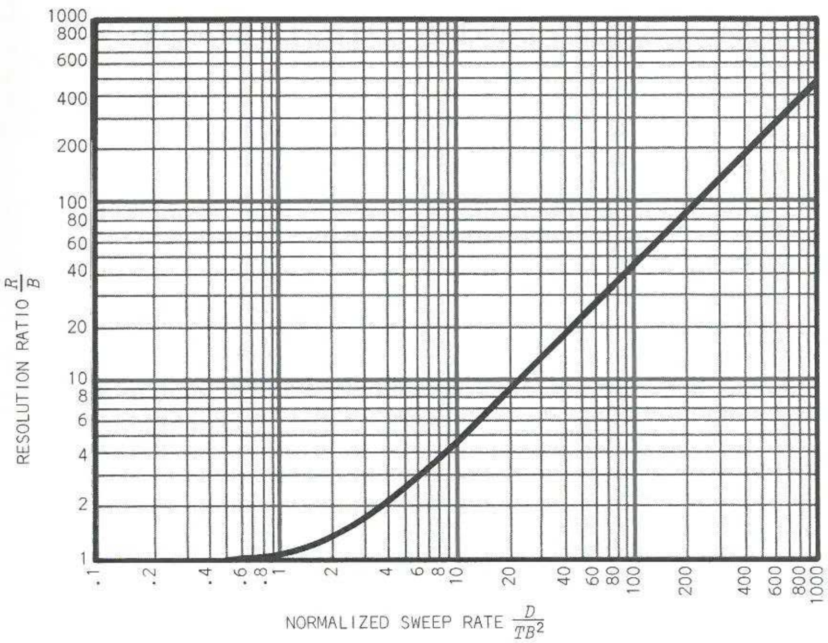


Fig. A-4. Loss in resolution as a function of normalized sweep rate. D = dispersion (Hz), R = resolution (Hz), B = bandwidth (Hz), T = time (s). Based on a Gaussian amplitude response.

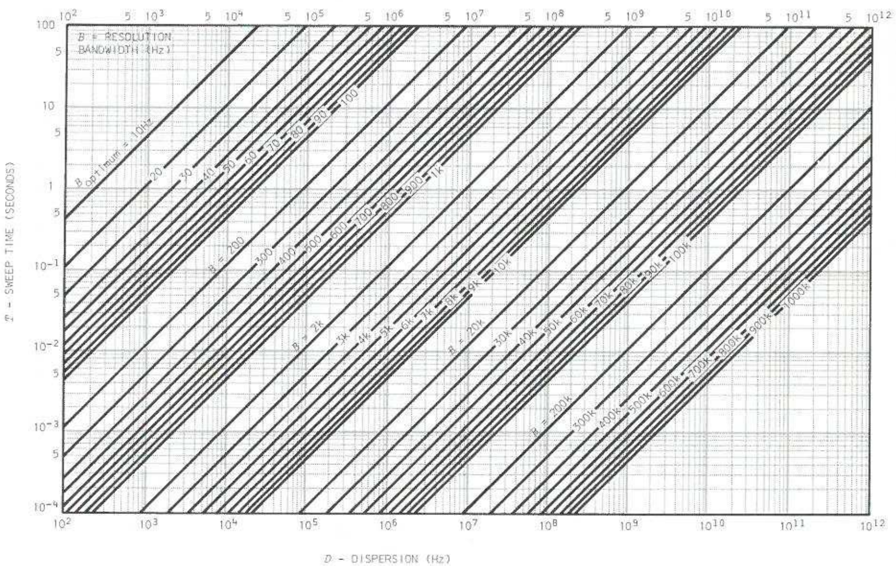


Fig. A-5. Optimum resolution setting for spectrum analyzers. Read values of B_{optimum} for a given dispersion and sweep time.

The concepts of optimum resolution and optimum resolution bandwidth are discussed in Chapter 5. Basically, at a fixed dispersion and sweep time there is one resolution-bandwidth setting which yields the narrowest resolution shape on the CRT display. These are called optimum resolution bandwidth and resolution respectively. Fig. A-5 is a graph showing the optimum resolution bandwidth (B_o) as a function of sweep time (T) and dispersion (D). The optimum resolution (R_o), which is the closest spacing of two signals that can be separated on the CRT screen, is related to optimum resolution bandwidth by

$$R_o = \sqrt{2}\, B_o \qquad \text{(A-12)}$$

It should be noted that Fig. A-5 is based on the assumption of a Gaussian amplitude response for the resolution amplifier.

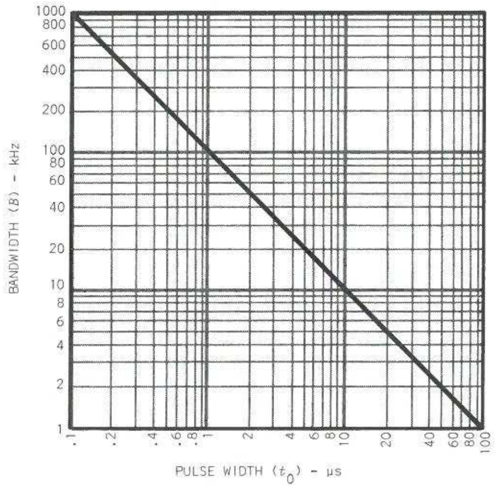


Fig. A-6. Resolution Bandwidth setting for pulsed RF computed from $Bt_0 = 0.1$.

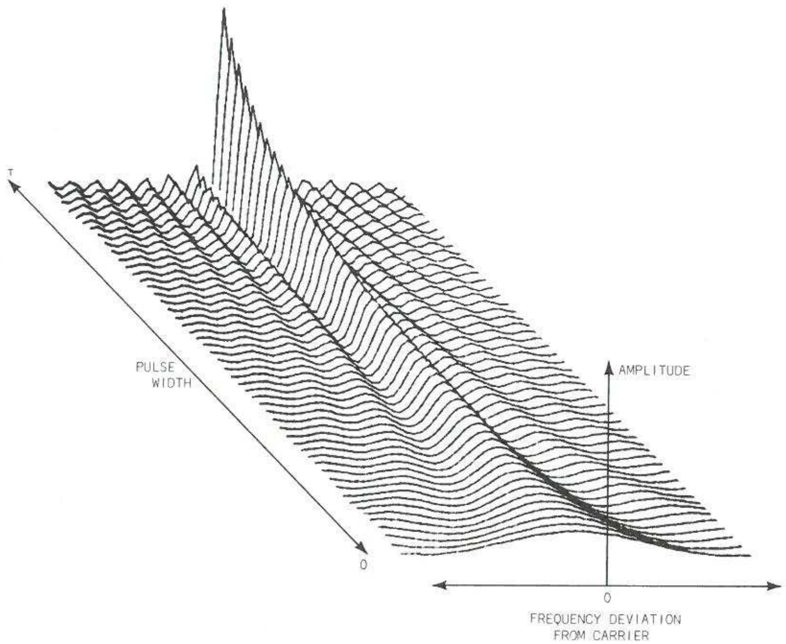


Fig. A-7. Pulse-width sidelobe-width pulse-sensitivity relationship.

PULSED RF

The details on pulsed-RF measurements are discussed in Chapter 9.

In order to display the fine details of the spectrum, it is necessary that:

$$t_0 B \leqslant 0.1, \qquad \text{(A-13)}$$

where t_0 is pulse width and B is resolution bandwidth. This relationship is plotted in Fig. A-6. When the pulsewidth-bandwidth product is greater than one tenth, spectrum shape details may be lost. As the pulsewidth-bandwidth product gets smaller, there is a progressive loss in sensitivity compared to a CW signal. Thus, $t_0 B = 0.1$ is the ideal setting for pulsed RF.

The loss in sensitivity for pulsed RF compared to CW can be computed from:

$$\alpha_{\mathrm{dB}} = 20 \log \frac{3}{2}\, t_0 B \qquad \text{(A-14)}$$

The loss in sensitivity occurs because, as the pulse width gets narrower, the energy spreads out over a wider frequency range. This is implied in the relationship between pulse width (t_0) and the frequency width of spectrum nulls (ΔF), namely,

$$\Delta F = \frac{1}{t_0} \qquad \text{(A-15)}$$

Fig. A-7 is a three-dimensional representation of the pulse width, sidelobe-frequency width and display-amplitude relationship. Fig. A-8 is a graph of equation (A-14), showing pulsed-RF loss in sensitivity as compared to a CW signal.

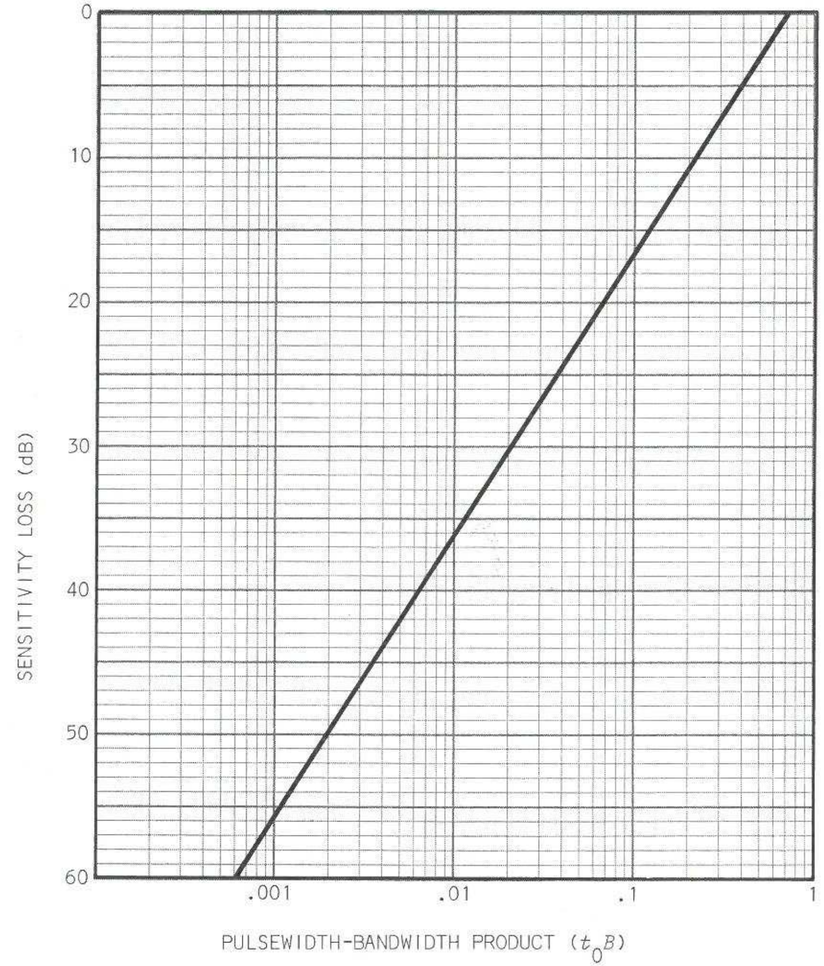


Fig. A-8. Loss in sensitivity, pulsed RF versus CW.

SYMBOLS

		first appears in chapter
a	instantaneous amplitude	4
$a_0, a_1 \ldots a_{m,\,n}$	Fourier-series amplitude constants	2
A	amplitude of waveform	2
$b_1 \ldots b_{m,\,n}$	Fourier-series amplitude constants	2
B	amplitude of waveform,	2
	filter bandwidth (Hz)	1
B_o	optimum resolution bandwidth (Hz)	5
c	speed of electromagnetic radiation	10
C	capacitance (F)	2
C_n	combined amplitude of nth harmonic in Fourier series	3
d	symbol for differentiation	2
d_n	amplitude of nth harmonic in complex notation of Fourier series	3
D	dispersion (Hz)	5
f	frequency (Hz),	1
	modulation frequency (Hz)	4
f_d	dial frequency (Hz)	6
f_D	Doppler difference frequency	10
f_i	image input frequency (Hz)	6
f_{IF}	IF amplifier center frequency (Hz)	1
f_{LO}	local-oscillator frequency (Hz)	1
f_{mo}	mixer output frequency (Hz)	6
f_{RF}	signal input frequency (Hz)	1
f_t	transmitted frequency (Hz)	10
f_0	center frequency (Hz)	1
$f(\,)$	function of ( )	3
F	carrier frequency	1
$F(\omega)$	Fourier transform	3
j	$\sqrt{-1}$	2

$J_p(t)$	Bessel function of the first kind of order p and argument t	4
k	Boltzman's constant	5
K	amplitude ratio in AM	4
L	inductance (H)	2
m	degree of modulation, harmonic number	4 2
n	harmonic number	2
N	noise power (W)	5
Q	charge (C)	2
R	resistance (Ω), resolution observed on CRT (Hz)	2 5
R_0	optimum resolution observed on CRT (Hz)	5
$\mathrm{Si}(x)$	sine integral of x	3
t	time duration (s)	2
t_0	pulse width (s)	3
T	absolute temperature (°K), period of waveform (s), sweep time per full CRT width (s)	5 2 1
v	velocity	10
x	unknown variable, variable angle (rad)	2 3
y	unknown variable	2
Z	complex number	2
Z^*	conjugate of Z	2
α	constant phase angle (rad), loss factor	2 5
β	modulation index	8
$\delta(t)$	impulse function	3
∂	symbol for partial derivative	2
ΔF	bandwidth, frequency deviation	3 4

ϵ	base of natural logarithms, 2.718...	2
θ	variable phase angle (rad)	2
π	3.141....	2
ρ	amount of error	2
$\sum_{n}^{m}$	summation between the limits n to m	2
τ	system response pulse-width for CW input (s), time shift interval (s)	1 3
ϕ_n	phase of nth harmonic in Fourier series (rad)	3
ω	radian or angular frequency, angular velocity	2
$=$	equals	1
$\neq$	is not equal to	2
$\approx$	is approximately equal to	4
$>$	is greater than	3
$\geqslant$	is greater than or equal to	3
$<$	is less than	2
$\leqslant$	is less than or equal to	5
$\rightarrow$	approaches, leads to, results in	3 9
$\leftrightarrow$	is equivalent by transformation	5
$\uparrow$	increase	6
$\downarrow$	decrease	6
$\sqrt{\ }$	square root	2
∞	infinity	3
$\lvert\ \rvert$	absolute value of quantity within the bars	10
$\lim_{x\to 0}$	indicating the limit of a term as x approaches 0.	3
$\int_a^b$	definite integral between limits a and b	2

BIBLIOGRAPHY

The author took cognizance of many references in the course of preparing this volume. Many of these are listed in the following. While no attempt has been made to give credit to the original developers of the material presented, one would be remiss in not specifically mentioning the influence of Goldman's *Frequency Analysis Modulation and Noise.* All who wish a better understanding of the physical interpretation of the modulation process should read this book.

The bibliography consists of two sets of references. One set of references is specifically referred to in the text. These usually contain a proof, analysis or other material which relates directly to a specific statement in the text. The other set of references, though perhaps of a more general nature, are no less, and sometimes even more, important to a complete understanding of the subject matter at hand.

CHAPTER 1

Engelson, *Spectrum Analyzer Circuits*, Tektronix Circuit Concepts Series, No. 062-1055-00, 1969.

Feigenbaum, "Introduction to Spectrum Analyzers," *Microwaves,* April, 1963.

Myer, *Getting Acquainted with Spectrum Analyzers*, Tektronix, No. A-2273-2.

CHAPTER 2

Goldman, *Frequency Analysis Modulation and Noise*, McGraw-Hill, 1948.

Lathi, *An Introduction to Random Signals and Communication Theory*, Scranton International, 1968.

footnote 1

Whittaker & Watson, *Modern Analysis*, Cambridge University Press, 1958.

footnote 6 Marcus, R. B., "The Significance of Negative Frequencies in Spectrum Analysis," *IEEE Trans*, EMC, Dec., 1967.

footnote 7 Cushman, R. W., "Spectrum Analyzer Myths," Report No. NADC-EL-6452, Feb., 1965.

footnote 8 Weber, *Linear Transient Analysis*, Wiley & Sons, Vol. II, 1954.

footnote 9 Weber, *Linear Transient Analysis*, Wiley & Sons, Vol. I, 1954.

CHAPTER 3

Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, 1962.

Reference Data for Radio Engineers, ITT, Fourth Edition, 1956, Chapter 35.

Stuart, *An Introduction to Fourier Analysis*, Methuen & Co., 1961.

footnote 1 Guillemin, *The Mathematics of Circuit Analysis*, Wiley & Sons, 1949.

CHAPTER 4

Black, *Modulating Theory*, Van Nostrand, 1953.

Goldman, *Frequency Analysis Modulation and Noise*, McGraw-Hill, 1948.

footnote 1 Whittaker & Watson, *Modern Analysis*, Cambridge University Press, 1958.

footnote 2 Jahnke & Emde, *Tables of Functions*, Dover Publications, 1945.

footnote 3 Giacoletto, "Generalized Theory of Multitone Amplitude and Frequency Modulation," *Proc IRE*, July, 1947.

footnote 4 Harvey, et al., "The Component Theory of Calculating Radio Spectra with Special Reference to FM," *Proc IRE*, June, 1951.

CHAPTER 5

Engelson, *Spectrum Analyzer Circuits*, Tektronix Circuit Concepts Series, No. 062-1055-00, 1969.

Feigenbaum, "Introduction to Spectrum Analyzers," *Microwaves*, April, 1963.

Lathi, *An Introduction to Random Signals and Communication Theory*, Scranton International, 1968.

footnote 1 Batten, et al., "The Response of a Panoramic Receiver to CW and Pulsed Signals," *Proc IRE*, June, 1954.

Chang, "On the Filter Problem of the Power Spectrum Analyzer," *Proc IRE*, August, 1954.

footnote 2 *Spectrum Analyzer Techniques Handbook*, Polarad Electronics Corp.

"Spectrum Analysis," *Application Note 63*, Hewlett-Packard Co.

footnote 3 Kicheloe, W. R., Jr., "The Measurement of Frequency with Scanning Spectrum Analyzers," Report SEC-62-098, Stanford Electronics Laboratories.

footnote 4 Engelson & Long, "Optimizing Spectrum Analyzer Resolution," *Microwaves*, Dec., 1965.

footnote 5 Montgomery, "Technique of Microwave Measurements," *Radiation Laboratory Series*, McGraw-Hill or Boston Technical Publishers, Vol. XI.

footnote 6 Metcalf, Von Allmen & Caprio, "Investigation of Spectrum Signature Instrumentation," *IEEE Trans*, EMC-7, No. 2, June, 1965.

CHAPTER 7

Engelson, "Interpreting Spectrum Analyzer Displays," *Microwaves*, Jan., 1966.

Goldman, *Frequency Analysis Modulation and Noise*, McGraw-Hill, 1948.

Langford & Smith, *Radiotron Designers Handbook*, RCA, 1953, Chapter 19.

CHAPTER 8

Engelson & Breaker, "Interpreting Incidental FM Specifications," *Frequency Technology*, Feb., 1969.

Engelson, "Interpreting Spectrum Analyzer Displays," *Microwaves*, Jan., 1966.

Goldman, *Frequency Analysis Modulation and Noise*, McGraw-Hill, 1948.

footnote 3 Charest, C. N., "Measuring Wide-Bandwidth FM Deviation," *EDN*, March 1, 1969.

footnote 4 Giacoletto, "Generalized Theory of Multitone Amplitude and Frequency Modulation," *Proc IRE*, July, 1947.

CHAPTER 9 Engelson, M., "Interpreting Pulsed-RF Spectra," *Microwaves*, March, 1969.
Engelson, "Interpreting Spectrum Analyzer Displays," *Microwaves*, Jan., 1966.

footnote 1 Montgomery, "Technique of Microwave Measurements," *Radiation Laboratory Series*, McGraw-Hill or Boston Technical Publishers, Vol. XI.
footnote 2 Engelson & Breaker, "Spectrum Analysis of FM'ing Pulses," *Microwave Journal*, June, 1969.

CHAPTER 10 Engelson, I., "Pinning Down Frequency Stability," *EDN*, May 15, 1969.
Frisch & Engelson, "How to get More Out of Your Spectrum Analyzer," *Microwaves*, May, 1963.
Shields, "Review of the Specifications and Measurement of Short-term Stability," *Microwave Journal*, June, 1969.

footnote 1 Davenport & Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, 1958.
Blackman & Tukey, *The Measurement of Power Spectra*, Dover Publications, 1958.
footnote 2 Peterson, "Response of Peak Voltmeters to Random Noise," *GR Experimenter*, Vol. 31, No. 7, Dec., 1956.
"Useful Formulas, Tables, and Curves for Random Noise," GR *Instrument Notes*, IN-103, 1967.
footnote 3 Metcalf, Von Allmen & Caprio, "Investigation of Spectrum Signature Instrumentation," *IEEE Trans*, EMC-7, No. 2, June, 1965.
footnote 4 Kauffman & Engelson, "Frequency Domain Stability Measurements," *The Microwave Journal*, May, 1967.

APPENDIX Abramowitz & Stegun, *Handbook of Mathematical Functions*, Dover Publications.
British Association for the Advancement of Science Mathematical Tables, University Press, Vols. VI & X.

INDEX

Absolute frequency, 134
AM/FM measurement, 186
AM/FM spectrum, 102
Amplitude loss factor, 115
Amplitude measurement, 135, 143
Amplitude modulation, 157
Angle modulation, 91
Angular velocity, 25
Area of impulse, 49, 70
Attenuator, calibrated, 135
Bandwidth relationships, 114
Baseline shift, 141
Bessell functions, 23, 95, 167, 169, 175, 267
  graph, 169
  partial table, 175
  tables, 267
  theory, 95
Bibliography, 285
Carrier null method, FM, 175
Carrier null table, 176
Center frequency shift by modulation, 66
Circular function, 25
Complex notation, 29, 39
Conjugate, 46
Conjugates, complex, 39
Continuous spectrum, 14, 58, 61
Convolution, 114, 125
  graphic illustration, 126
  theorem, 75
CW sensitivity vs sweep rate, graph, (Appendix), 272
Cyclical (definition), 17
dB relationships, (Appendix), 263
Definition of terms, 245
Degree of modulation, 89
Dense spectra, 14, 61, 118, 191
Dense vs line spectrum, 210
Deviation, FM, 168
Deviation linearity measurement, 181
Dispersion, 8
Display, 159, 177, 208-209, 225-226, 230, 262
  definition, 209
  filter bandpass characteristics, 226
  frequency response, 230
  optimum resolution, 262
  single-tone AM, 159
  sweep time vs spectrum (definition), 208
  wideband FM deviation, 177
  zener noise, 225
Distortion measurements, 224
Doppler velocity measurement, 239
Drift measurements, 236
Dynamic range, useful, 144
Effect of control settings, 209
Effect of measurement time, 257
Effect of pulse shape, 196
EMI measurements, 233
Energy distribution, 117, 121, 122
Energy per unit bandwidth, 63
Energy, sideband, 89
Example: carrier null, 176
Example: derive one spectrum from another, 77
Example: deviation linearity, 181
Example: deviation measurement, 108
Example: percent AM modulation, 108

Example: rectangular pulse
analysis, 81
Example: Spectrum Analyzer frequency
relationships, 15
$f(\omega)$, $F(\omega)$, frequency domain, 74
$f(t)$, $F(t)$, time domain, 74
Feedthrough, IF, 138, 141
Filter characteristics, display, 227
First-sideband null, 182
FM energy distribution, 105
FM spectrum, generation, 104
Fourier analysis, 43, 67
Fourier applications, 47
Fourier coefficients, 41, 45, 67
Fourier expansion, 32
Fourier integral, 63
Fourier Series, 32
truncated, 40
Fourier transform, 64-65, 269
chart, (Appendix), 269
examples, 77
pair, 64
table, 74
Frequency, absolute, 134
Frequency axis, 8
Frequency conversion, 137
Frequency deviation, 93
Frequency domain, 5, 34, 71, 88,
134, 158
AM, 88, 158
illustration, 34
impulse, 71
measurement, 134
Frequency linearity, 258
Frequency modulation, 97, 98-100,
102, 105, 167
energy, 100
sidebands, 102
spectrum, 99, 105
Frequency scale, 15
Frequency separation, 7
Frequency/time-position, 10
Frequency translation, 111
Gain suppression, 145
Gaussian response, 115, 117
Gibbs phenomenon, 56
Harmonic amplitudes, 67
Harmonic conversions, 140
Harmonic distortion, 148
Harmonic zeros, 59, 62
Harmonics in squarewave, 53
IEEE definition, 250
IF feedthrough, 141
Image, 16
Image frequency response, 142
Image identification, 153
Image response, 138
Imaginary numbers, 29
Impulse approximations, 72
Impulse energy distribution, 73
Impulse, frequency domain, 71
Impulse function, 48, 70
Impulse, spectral distribution, 71
Impulse strength, 49, 70
Incidental FM, 102, 237
Infinite harmonics, 61
Initial phase, 86
Instantaneous frequency, 91
Instantaneous power, 86
Intensification effects, 187, 189
Intermodulation distortion, 149, 164
Least squared error, 33
Linear time-invariant, 34
Line spectra, 14
Lobe, main, 66, 81, 120
Lobe spacing, 82
Lobe width, 82
Long pulse; CW, 120
Low-level modulation, 160
Magnetic shielding test, 233
Manual sweep, 243
Measurement interval, 118
Measurement limitations, 135
Measurement methods, explanation, 253
Miscellaneous applications, 219
Missing pulses determination, 202
Modulation, 85
Modulation index, 99, 168, 173
Modulation percent, AM, 158
Modulation rate, determining, 185
Modulator on/off ratio, 204
Multitone FM measurement, 188
Multitone modulation, 161
Narrowband FM, 169, 170
Network response, 107
Network responses, sum, 54
Networks, physically realizable, 57
Noise power, thermal, 123
Noise vs bandwidth, 123
Optimum resolution, 116, 124
Optimum resolution setting, graph,
(Appendix), 276
Orthogonal functions, 20
set, 68
table, 38
Overshoot, 56
Period, 17, 28
Periodic (definition), 17
Periodic waveform illustration, 18
Phase angle, initial, 28
Phase deviation, 92
Phase-frequency relations, 94
Phase modulation, 92
Phase response, 115
Power, sideband, 89
Pulsed RF, 14, 83, 191, 193, 198,
213, 215, 278
display intensification, 215
displays, 193
effects of FM, 198
fine detail, 213
resolution bandwidth graph, 278
sensitivity and dynamic range, 213
Pulsed RF vs CW, loss in sensitivity, 280
Pulsed-RF wave, 66
Pulsed signals, 117
Pulsed signal sensitivity, 124
Pulsed signal time-frequency diagram, 119
Pulse shape, effect of, 196
Pulse train, 50, 119
Pulse train, narrow, 58
Pulse train, rectangular, 60
Pulse vs CW, 120, 121
Pulse width, 10, 13, 112
Pulse width parameters, 114
Q, determination, 227
Radar performance, 194
Radian frequency, 25
Random-noise measurement, 223
Reciprocal spreading, 73
Rectangular pulse analysis, 81
Relative amplitude, 143, 146
Relative frequency, 134
Repetition rate lines, 120, 191
Resolution, 8
Resolution bandwidth, 106, 116, 191,
275
dense spectrum, 191
graph of, 275
Resolution calculation, 128
Resolution distortion, 115
Response, CW, 112
Response, DC and phase, 44, 56
Response skirts, 217
Response, true, 155
Sensitivity, 123
Sensitivity, calculation, 127
Sensitivity for pulsed RF, 192
Sensitivity vs bandwidth and noise, 271
Sidebands, 88
Signal frequency identification, 153
Signal interpretation, 133
Signal purity measurements, 236
Signal types, 14
Sine function, 23
Sine integral, 79
Sine-squared pulse, 76, 198
Sinewave generation, 26
Sinewave properties, 23
Sinewaves, infinite duration, 49
Single sideband, 91
Single sideband AM, 162

Sinusoid, 86
Sin x/x, 52, 61, 67, 81
Si, sine integral, 79
Skirt selectivity, 145-148
Skirt selectivity effects, 216
Slope detection, 170
Small signal in presence of large,
144, 146
Spectral density, 63
Spectral lines, 36
Spectrum analyzer as an indicator, 227
basic, 113
configurations, 7
(defined), 5
Spectrum, continuous, 58
Spectrum null, 59
Spectrum shape, 191
Spurious display, 155
Spurious response, 16, 137, 141
Squarewave configurations, 19
Squarewave synthesis calculations, 219
Sum of two sawtooth waves, 55
Superposition, 83, 103
Superposition in Fourier, 54
Suppressed carrier, 90
Suppressed carrier AM, 162
Sweeping signal, 13
Sweeping-signal spectrum, 110
Swept-frequency converter, 228
Swept local oscillator, 13
Symbols used in text, (Appendix), 281
Synchronized sweeper, 228
System resolution, 11
T/t_0, 59, 62
Telemetry subcarrier tests, 234
Time axis, 8
Time domain, 5
Time domain, AM, 87, 157
Time-domain unit impulse, 70
Time-frequency concept, 5
Time-frequency diagram, 10, 113, 142
Time position, 112
Time-variable network, 42
Transducers, 240
Transient response, 35, 106
Trapezoid pulse, 196
Triangular pulse, 197
Trigonometric functions, 30
Trigonometric notation, 30
Ultrawideband FM, 183
Useful dynamic range, 144
Vector, rotating, 28
Vestigial sideband, 91
Vibration measurements, 240
Waveform analysis, 219
Wideband FM, 175
Zener noise, 224

NOTES